DOCUMENT RESUME

ED 276 982 CS 008 623

AUTHOR Watson, Alan J.; And Others
TITLE Cognitive Development and Learning To Read: A Five Year Longitudinal Study.
PUB DATE [84]
NOTE 136p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC06 Plus Postage.
DESCRIPTORS Beginning Reading; *Cognitive Development; *Developmental Continuity; Developmental Stages; Elementary Education; Foreign Countries; *Learning Processes; Longitudinal Studies; Multivariate Analysis; Oral Language; Piagetian Theory; *Reading Processes; Reading Research; *Reading Skills

ABSTRACT

A cognitive developmental theory of reading was proposed and tested in a longitudinal study that (1) compared the development of conceptual reasoning (Piaget's "operativity") and oral language (vocabulary and grammar closure) of 148 Australian children during the first 5 years of school and (2) traced the relationship of these variables to reading during that time through multitrait/multitime analysis. Results indicated that the development of operativity is surgent and discontinuous while oral language is stable and continuous; the influence of language awareness, IQ, or reading instruction could not account for these developmental characteristics. Findings suggested that the operativity measures of this study are tapping a powerful developmental influence that is partly distinct from oral language in its effects on reading. Analysis indicated that this influence was part of a broad spectrum of change in mental processing that occurs during middle childhood. Emerging conceptual capacity was found to be a necessary although not sufficient condition for the development of reading ability. As well as exerting an immediate influence, findings showed that operativity produces a pervasive delayed effect in its causal relationship with reading comprehension that, during the early school years, slowly takes on the decision-making, conceptual reasoning character of that operational influence. (JD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

ED276982

COGNITIVE DEVELOPMENT AND LEARNING TO READ: A FIVE YEAR LONGITUDINAL STUDY

ALAN J. WATSON
ST. GEORGE INSTITUTE OF EDUCATION
SYDNEY COLLEGE OF ADVANCED EDUCATION

KENNETH C. SINCLAIR
DEPARTMENT OF EDUCATION
UNIVERSITY OF SYDNEY

HERBERT W. MARSH
DEPARTMENT OF EDUCATION
UNIVERSITY OF SYDNEY

GEORGE H. COONEY
DEPARTMENT OF BEHAVIORAL SCIENCES
MACQUARIE UNIVERSITY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

A. J. Watson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

CS 008 623

COGNITIVE DEVELOPMENT AND LEARNING TO READ:

A FIVE YEAR LONGITUDINAL STUDY

INDEX

READING AND THE DEVELOPMENT OF COGNITION 1

A Range of Evidence 2
Operational Thought and Reading 4

Conservation 5
Classification and Seriation 7
Perceptual Exploration 9

Operational Thought and Reading - Explaining the Link 14

THE PURPOSES OF THE STUDY 18

Developmental Nature of the Constructs 18
Relationship of Operativity to Reading 25

BRINGING THEORY AND METHOD TOGETHER 27

Measuring Developmental Capacities 28
Construct Validity: Multitrait-Multitime Analysis 32
Validity of Variable Groupings 36
Individual Differences and Qualitative Change 38
Hypotheses of the Study 41

METHOD

Design 42
Sample 42
Materials 43

Reading Measures 43
Piagetian Measures 44
Language Measures 46

Procedures 46
Data Analysis 48

RESULTS

Effects of Repeated Testing 51
Effects of Historic Time Period 52
Establishing Construct Validity 53

Validity of Predictor Variables 53
Validity of Criterion Variables 55
Validity of Variable Groupings 56

Developmental Change 62

Development of Predictor Variables 62
Development and Construct Validity 68
Development within the Criterion Variables 69

Relationship of Predictor to Criterion Variables 71

Prediction of Dependent Variables 72
Pattern of Effects 77

Summary of Results 84

DISCUSSION 88

A Developmental Contrast: Operativity and Oral Language 88
Stage Progression 90
The Relationship of Operativity and Reading 93
Some Current Issues 96

Critical Level 96
Prediction over Time 99

CONCLUSION 101

ABSTRACT

On the basis of a range of theory and research it is argued that reading is fundamentally a conceptual reasoning task. Studies of strategic processes in reading, the examination of children's concepts of the nature of print and the naturalistic study of oral reading behaviour, give support for this contention. Since, in Piagetian theory, the emergence of concrete operations gives major new impetus to the conceptual reasoning capacities, the significant positive relationship of concrete operations to reading provides further evidence for this view. This relationship also gives a basis for studying the development of these conceptual processes in regard to the emerging reading competence of young school pupils. A cognitive developmental theory of reading is proposed and is tested in a longitudinal study which compares the development of conceptual reasoning (Piaget's "operativity") and oral language (vocabulary and grammatic closure) of 148 boys and girls during the first five years of school and traces the relationship of these variables to reading during that time. The construct validity of the variables and of their grouping is assessed principally by means of multitrait-multitime analysis - a technique developed for the study from multitrait-multimethod analysis (Campbell and Fiske, 1959). The results support the construct validity of the variables and their grouping and suggest that conceptual reasoning and oral language display different forms of developmental change over this period. The development of operativity is found to be more surgent and discontinuous while oral language is more stable and continuous. This difference is manifest in their different relationships with reading as revealed by path analysis.

Operativity exerts both direct and indirect effects on reading which are distinct from the effects of oral language at least in this middle class sample and with the measures employed. These effects do not seem to be explained by the influence of language awareness, IQ or reading instruction. The nature of developmental change in reading related capacities is discussed as are its implications for learning to read. Issues in the literature concerning a critical developmental level for learning to read and a sleeper effect of operativity on reading are considered.

Reading and the Development of Cognition

In recognition that reading involves "many of the most intricate workings of the human mind" (Huey, 1908), many recent theoretical models of learning to read have sought to explain the nature of the learner's internal processing. Psycholinguistic views of Goodman (1970), and Smith (1971) argue that reading is a decision making activity in which minimal language cues are utilised to determine meaning. Information processing models typically describe the internal mechanisms of attention, perception and memory called upon during reading (e.g., La Berge and Samuels, 1976). The present study proposes a cognitive theory of learning to read in which it is argued that the decision-making processes of the psycholinguistic model and the co-ordination of activity implicit in information processing views demand an explanation in terms of higher mental processes. It is contended that any explanation of learning to read which overlooks the executive functions of the mind fails to take account of the central and most critical activities of the learner.

It is further argued that, since reading is not a unitary task learnt in a short time but a complex integration of many learnings over a lengthy period of time and a period during which young children are usually experiencing major changes in their mental powers, an explanation of learning to read in terms of the child's development is called for. The child learns one task (or skill, or understanding, or piece of knowledge) and is thereby enabled to progress to the next. A reading curriculum statement expresses this well:

> Progress is marked by leaps, pauses, periods of consolidation, regression and by periods of slow or moderate progress. It is a succession of merging and overlapping phases rather than a series of isolated skills taught in rapid succession. (Department of Education, 1979, p.7)

This makes the task of explaining its nature particularly complex. Reading may well call upon different capacities at different phases of its learning. The ability of children to respond to reading instruction may well be subject to differences in the general developmental status of their cognitive systems.

Range of Evidence

A wide range of evidence from diverse theoretical and practical perspectives provides firm support for a cognitive developmental view of reading. Studies of the strategy usage of good and poor readers have demonstrated - both at the level of letter synthesis to form syllables and words (Katz, 1977; Scheerer-Neuman, 1977-1978 cited in Valtin, 1978) and at the level of word integration to form phrases and extended discourse (Denner, 1970; Keeton, 1977; Steiner, Wiener and Cromer, 1971) - that good readers impose structure upon print in a significantly more effective way than do poor readers (Golinkoff, 1976; Ryan, 1979). More precise specification of this strategic operation is suggested by studies of cognitive style in good and poor readers. Poor readers were unable to handle hypotheses and inferences and lacked the conceptual reasoning capacities possessed by good readers (Serafica and Sigel, 1970).

Furthermore, the provision of a conceptual structure to deal with text has been demonstrated, in several experimental studies, to improve the comprehension of poor readers. The range of strategies taught includes top-level structure (Meyer, 1984) semantic mapping (Sinatra, Berg and Stahl-Gemake, 1984) hypothesis testing (Samuels, Dahl and Archwamety, 1974), active comprehension (Singer & Donlon, 1982) and visual imagery (Levin, 1973). These seem to strengthen the organizing, conceptual activity of the reader and are consistent with the cognitive explanation of reading proposed.

Another strand of evidence which supports the view proposed and clarifies the nature of the cognitive components of early reading is the study of the young child's concepts of print. The beginning reader's concepts of the nature and purpose of print have shown a consistent and positive relationship to progress in reading. (Ehri, 1979; Francis, 1973) Metacognitive insight into the task seems to be an important aspect of the conceptual reasoning required as a child learns to read. (Watson, 1984)

Further evidence for this view may be found in the naturalistic study of oral reading. This has led to the description of reading as a psycholinguistic decision-making process (Goodman, 1969) or as a perceptual search-and-check operation (Clay, 1972) in which language cueing systems are used to determine meaning. Although this seems to imply a conceptual reasoning activity, it has been explained as language learning (Goodman, 1969) or as a perceptual task (Clay, 1972) because, it seems, in each case, the automatic processing of reading is believed to be too fast for thought to be its basis. This is an important objection to the cognitive theory proposed here since competent reading clearly does involve a predominance of automatic activity in which thinking plays no apparent part. However, even the most competent reader must occasionally stop to puzzle over and reason about the meaning of a word or passage and certainly must possess that reasoning ability irrespective of how frequently it is called upon. The learning-to-read process is thus likely to be a more revealing place to look for knowledge of the underlying nature of reading and here the evidence gives strong support for the critical importance of conceptual reasoning for reading.

Further explication of the relationships of the automatic and higher order processes in reading is suggested by the reviews of White (1965, 1970) in

which he argues that there are two response levels, an automatic or associative level and a decision or cognitive level, which have different developmental histories. The associative level, which can claim responses learnt at the cognitive level, is well established in early childhood but the cognitive level begins or develops markedly in the 5 to 7 age period. Reading seems to be a prime example of this two layered response capacity with much learning initially taking place in the decision zone, being claimed and used by the associative zone but able to be checked at the decision zone as necessary.

It is at about the same time a child learns to read, that White (1965, 1970) notes there is a beginning of, or a marked upsurge in the capacity to use the cognitive level of response, and also during the 5 to 7 age period that Piaget claims there is a major qualitative change in the child's capacity for conceptual reasoning, the concrete operational stage.

Operational Thought and Reading

Operational thought in Piaget's theory is a major landmark in the emerging human capacity to make sense of the world because it denotes the emergence of a far more powerful and efficient conceptual process in the human life cycle. In this view reading is highly reliant on the stage related cognitive development of the young child.

For Piaget a concept is an internalised action. In early childhood concepts derive from sensory-motor actions while in middle childhood they are also based upon logical groupings or mental operations. The child of 6 or 7 learning to read is passing from "sensory-motor intelligence" to "conceptual intelligence". The difference lies in the capacity for reversible operations which allow a simultaneous representation of experience and so free the child

from the limitations of space and time. It may not be by chance that children show an ability to learn to read at 6 or 7 years, the time when operational thought emerges. If reading is a conceptual activity as the literature has suggested, there is a need to clarify the conceptual foundations upon which it is built. If Piaget's view of cognitive development in middle childhood has some validity we may well expect the operational capacity to relate to reading progress. Studies which have examined the three infralogical operations, conservation, classification and seriation and a related capacity, perceptual exploration, in their relationship to reading, will be reviewed.

Conservation. Of the studies that have examined concrete operations and reading (or reading readiness), most research has focussed on conservation. Waller (1977), who drew together the results of some ten studies, found that overall, correlations were low to moderate ranging from .17 to .44 but most commonly from .25 to .35. Kaufman and Kaufman (1972), who found the highest correlation, studied 80 children whose conservation of length and number (inter alia) were measured in Kindergarten and whose reading (Stanford Achievement Test) was measured 16 months later in first grade. To assess conservation of length the children were asked to compare lengths of straight and curved pieces of Play Doh and to explain the reasons for their answers. Conservation of number was measured by asking the child to demonstrate knowledge of rudimentary adding and subtracting (with objects), of identity conservation (whether the number of candies stays the same when poured from a plate to a glass), and of equivalence conservation (whether two equal rows of objects remain equal after one row undergoes shape transformation). Conservation of length correlated .30 with reading while conservation of number correlated .44. This latter correlation is the largest reported in the literature. However, close scrutiny shows that it is not based upon a valid measure of conservation since adding and subtracting involve the unit concept which requires a co-ordination of both classification and seriation.

In summarising the evidence, Waller (1977) concluded that with both males and females and across a range of social statuses

> there generally is at least a low positive correlation between performance on a variety of tests of reading and reading readiness on the one hand, and measures of level of cognitive development as regards concrete operations (specifically conservation) on the other (p.14).

In an important study not included in the Waller review, Lunzer, Dolan and Wilkinson (1976) administered a battery of 28 tests of operational thought, language and memory to a sample of 183 Kindergarten children of three different social levels; from privileged, intermediate and underprivileged areas. Twelve months later they administered two reading tests (Schonell Word Recognition, Neale Comprehension). The validity problem of the Kaufman and Kaufman study seems to have been overcome since the conservation of number test included 1 for 1 correspondence but no adding or subtracting. Conservation of length asked children to assess the equivalence of equal and unequal rods and pieces of string in parallel alignment, displaced or rearranged. The multiple correlation of all predictors with Schonell Word Reading, was .65 and with Neale Comprehension was .54. Although conservation of number and length correlated at moderate levels with reading (Schonell) (.31 and .25 respectively, E.A. Lunzer, personal communication, 1977), when conservation was dropped from the battery the multiple R changed very little. It was concluded, therefore, that conservation did not contribute significantly over and above other operative and learning measures towards the prediction of reading.

Thus, although there is some relationship between conservation and reading, it is only at a low to moderate level and the variance which it controls is also controlled by classification and seriation which overshadow

it. More research attention therefore ought to be directed towards classification and seriation in relation to reading.

Classification and Seriation. Studies of classification and reading have not been as numerous as have been studies of conservation and reading but the correlations reported generally have been somewhat higher, usually in the vicinity of .35 (Waller 1977). Kaufman and Kaufman (1972), for example, measured logical classification by two-way block sorting (colour and shape) class inclusion ("some" and "all") and multiple classification tasks. This correlated .35 with reading (Stanford Achievement Test) assessed 16 months later.

The battery of Piagetian classification measures of Lunzer, Dolan and Wilkinson (1976) included "all and some", class inclusion (unequal partition and nested classification) and intersecting classes (simultaneous classification upon the basis of two criteria). These correlated at levels of .43, .32 and .54 respectively with reading (Schonell). In classification of intersecting classes the child's incorrect responses were corrected and subsequent items allowed the child to profit by earlier experience. Lunzer, (in personal communication, 1977) writes "If you pressed me to comment upon the interaction task as a predictor, I would possibly venture that it combined operativity with a fair amount of learning. But one should be chary about making such fine grade distinctions post hoc!" In the multiple regression analysis Lunzer grouped classification with seriation upon the basis of an earlier factor analysis. Seriation comprised single seriation (e.g., ordering different sized rods) and multiple seriation (e.g., cylinders varying in shape and size arranged on a 3 x 3 matrix). When classification and seriation were omitted from the multiple correlation it dropped from .65 to .61 - the largest

single effect shown, and one which the covariance F test revealed is significant at the .01 level for the whole sample and for the privileged and under-privileged socio-economic groups. For the intermediate socio-economic level it did not attain significance. A comparison of the effects of classification and seriation shows that single and multiple seriation (which correlated .47 and .39 respectively with reading) were the best representatives of this part of the battery. Classifiation and seriation have both shown moderate to strong correlations with reading at grade one but this and other research indicates that seriation is perhaps the most promising of the three Piagetian predictors.

Scott (1970) compared results of single seriation given to kindergarten children with their April third grade reading scores as measured by the Iowa Test of Basic Skills. The 151 boy and girl subjects were from lower and middle class backgrounds and included blacks and whites. Seriation was measured with a thirty item test of various size and pattern ordering tasks, matching of two series and an assessment of trial and error versus operational modes of response. A correlation of .50, significant at the .01 level, was found. The relationship was stronger amongst white than amongst black children. Kaufman and Kaufman (1972) asked children to make staircases out of 5 then 9 slats (ranging from 2 to 6 inches long) and then to insert two additional slats into the larger staircase. They were finally asked to count the steps and tell how many steps a doll had climbed. Although this final step appears to confound seriation with classification (since it requires the unit concept) the results of Lunzer, Dolan and Wilkinson (1976) indicate that this may not be important since the two seem to call upon the same capacities (as shown by the factor analysis). This seriation task, when administered to 30 white upper-middle class kindergarten boys and girls, correlated almost as strongly with reading in first class (.55) as did the whole Piagetian battery

(.58) and was a stronger correlate of reading than I.Q. (.48, Lorge-Thorndike) or mental age (.51, Lorge-Thorndike).

Taken together these three studies provide impressive evidence for the existence of an important relationship between seriation and reading. Similar results have been obtained in two different English speaking countries, for both boys and girls, from under-privileged and privileged social groups, with independently constructed predictor measures and different criterion tests. The relationship extended from the predictor administered in Kindergarten to the criterion in one case some 18 months later while in another, some 30 months later. Certainly any investigation of the nature of conceptual reasoning as required by learning to read is well advised to include seriation.

Perceptual Exploration. Perceptual exploration, the developmental capacity to perceive a complex visual field, is a Piagetian concept which the evidence suggests is partly related in its development to operational thought and to success in learning to read. Since Piaget's views upon perception are little known it is necessary to consider them (and some supporting studies of Vurpillot and Zaporozhets) as they relate to conceptual ability before examining perception, reasoning and reading.

During early childhhood perception is "syncratic" or "global", i.e., the child lacks the ability to synthesise and analyse and so cannot co-ordinate wholes and parts (Piaget, 1969). This is the outcome of an insufficiency of exploratory activity. This activity directs eye movements and determines pauses or centrations during the examination of a figure. Exploration develops with age, requires practice and direction and is related to intelligence which in part directs its functioning. Perceptions and concepts each arise from sensory-motor activity. It has often been thought that concepts derive from perceptual activities. Piaget (1969) does not believe this is so. Perceptual

activity is never self-sufficient but is directed by sensory-motor intelligence or representational (operational) intelligence:

> perceptual activities exhibit characteristics of composition which are partially isomorphic with the forms of operational composition ... share common filiations with early stages of sensory-motor activities. (Piaget, 1969, p.311f)

There is no dualism of experience (i.e., perception) and deduction, he argues, since experience is a progressive structuring (not a recording) and deduction is a co-ordination of operations. Thus the difference between perception and reasoning is only in the degree of mental activity involved (Piaget, 1969).

In an authoritative recent statement upon perceptual development in children, Vurpillot (1976) reviews the evidence upon perceptual identification of wholes and parts. What is perceived, she concludes, will be determined by the perceptual organisation of the field and the cognitive structures of the perceiver, as Piaget has asserted. The capacity to see both parts and wholes in an ambiguous figure seems to depend upon the emerging cognitive operations.

> It is this double role of a single unit, a role analagous to that taken by a common element in related classes, that children appear to be incapable of appreciating before the age of about 7. Up to that age-level a detail of a figure can constitute either an individualised and identified structure or a part of another structure, also individualised and identified - but it cannot take both roles at the same time. (Vurpillot, 1976, p.147)

The age of 7, Vurpillot emphasises, is only an approximate indication, "a

mean around which the dispersion is very great." Demonstration of this capacity can be earlier (or later) depending upon such things as whether the task involves differentiation or identification, whether the contours are continuous or broken and upon how familiar and meaningful is the content. Zaporozhets' review of Soviet research into haptic perception (Zaporozhets, 1969) gives support for the developmental pattern proposed by Vurpillot and especially for the relationship of perception to mental processes. Haptic exploration begins to be more systematic and thorough at 5 to 6 years and at 6 to 7 years becomes adult-like with systematic exploration of all the contours of a shape.

> These findings would seem to confirm that it is only by age 5 that a system of strictly perceptual-cognitive functions begins to develop, the characteristic feature being extensive movements that are carried out consecutively by the receptor apparatus (Zaporozhets 1969, p.96).

This research supports the view that important aspects of perception gain significant impetus from the changes in mental function which take place during middle childhood. If perceptual exploration is "partially isomorphic" with the mental operations and is, to an important degree, directed by them, it would not be surprising if it contributed to the conceptual reasoning which the foregoing discussion has suggested is central to learning to read. Several empirical studies have suggested this.

A series of studies upon perceptual exploration (Piaget uses the term "activity", Elkind, "regulation") and reading have been conducted by Elkind and his co-workers (1975) who developed a set of pictorial items to monitor perceptual ambiguity (one figure capable of two interpretations) and perceptual integration (independent items forming a whole) in part replicating the work of Dworetzki (1939, cited in Vurpillot, 1976). In cross sectional and

experimental research with children aged four to eleven years there was an increase with age in ability to perform these tasks and to benefit from training. The results are interpreted to confirm the assertion of Piaget that perceptual development is a constructive process which shows critical change during middle childhood and is related to the structures of the intellect.

Working amongst children with the perceptual exploration tests Elkind was struck by the apparent relationship of perceptual capacity and reading competence. Investigation of this relationship showed that decentration of perception (going beyond the initially salient feature of a complex visual field) was significantly related to reading (Elkind, Horn and Schneider, 1965) and further research (Elkind, Larson and Van Doorninck, 1965) revealed that perceptual regulations act, in part, independently of general intelligence in this relationship. To test the hypothesis that children would improve in reading if given training in perceptual regulations, 60 inner-city second grade negro pupils were given instruction for three half hour lessons a week for 15 weeks in two groups matched for perceptual activity and reading achievement (Elkind and Deblinger, 1969). The control group read orally around the class from a basic reader and did vocabulary, grammar and comprehension exercises provided in the teacher manual. The experimental subjects who were trained with a set of non-spoken exercises to stimulate perceptual regulations, improved in their recognition of words and word forms significantly more than did the controls. Comprehension showed no significant difference between the groups.

Some caution is needed in interpreting this study. On the one hand the group instruction given to the controls does not represent a very efficient teaching pattern. No particular account of individual pupil reading needs was taken since pupils read aloud while others followed and general exercises were done from the board. An examination of the results bears out the inefficiency

of the methods. The control group's mean score improvement was only 0.57 from 15.79 to 16.26 in word recognition and 1.41 from 8.14 to 9.55 in comprehension which represents very little progress after a fifteen week course of lessons. On the other hand the experimental group's instruction may have provided some reading experience since it includes exercises such as using jumbled words and singular-plural pairs. The difference between the two groups could be explained by their different reading experiences rather than by the training in perceptual exploration. The case for a causal relationship between perceptual exploration and reading progress is yet to be established. A specific focus on subjects scoring low in regulations and in reading seems to be required and more effective teaching strategies and lesson patterns need to be used. Stricter control across the treatment groups over factors such as level of motivation and format of lessons can be exercised, despite the use of ecologically valid setting, the school classroom. However, Elkind's series of studies has added greatly to the developmental explication of perception and suggest that perceptual exploration, directed in part by operational thought, is a capacity called upon in learning to read.

Although there has been no independent replication of Elkind's work, support for his view of the perceptual-cognitive process in reading can be found in some work of Kolers (1975). Kolers initially demonstrated that it is possible to distinguish experimentally the processing of the graphemic aspects of a sentence from the interpreting of the message embodied in those graphemes. Next he compared matched groups of good and poor readers (with mean age 12.1) upon their ability to encode semantic and graphemic features of sentences. Contrary to some expectations, his results gave no support for the view that impaired reading is the result of a general impairment of language skills. The two groups did not differ in grammatical usage, in memory of what was read or in speed reading improvement. The good readers were better than the poor

readers "in their ability to analyse and remember graphemic patterns". Kolers remarks that poor readers' difficulty seems to have a perceptual basis but "at a more cognitive level of performance than is usually measured by tests of visual function....This vague middle ground clearly requires better specification" (p.290). There seems to be a remarkable parallel between this and Elkind's work on perceptual exploration in reading. It seems to be this same "vague middle ground" that Elkind is exploring.

The evidence reviewed in this section suggests that perceptual exploration is a perceptual-cognitive function related to success in reading. The precise relationship of perceptual exploration and operativity (as seriation) in regard to reading requires empirical investigation.

Operational Thought and Reading - explaining the link

The empirical evidence reviewed indicates that there is a significant relationship between operativity, especially seriation and perceptual exploration, and reading. It might well be wondered, however, how the processes involved in operativity, as measured through picture observation and object manipulation tasks, are linked to those involved in a language activity, reading? At first glance it seems to be a rather unlikely predictive combination.

An answer may begin to emerge in regard to perceptual exploration by considering the grapho-phonic cues children use as they recognise words. The general configuration of letters will obviously be an important cue for many words; "Elephant" and "rhinoceros", for example, have gross features which make them readily distinguished from "ant" and "flea". But there are limits to this strategy. Finer discriminations are required to distinguish "giant"

and "great", "house" and "home", and in this, perceptual decentration seems to play an important role.

The parts of words young children used in recognition were studied by Marchbanks and Levin (1965) with a sample of kindergarten and first grade pupils. It was found that the first letter was most important for word recognition, the last letter the next in importance and the letters in between were given least consideration. It seems that children whose perceptual exploration is little developed are more likely to be held by the salient aspects of the word, e.g., the first letter, and will take account of less dominant parts of it with more difficulty. In a similar way children's perceptual exploration seems to be implicated in their processing of words in a sentence. The child whose perception is centred on the salient words or phrases in a passage, e.g., the first words or best known words, is likely to have difficulty allowing each word to take the weight it was intended to have by the meaning of the sentence.

A further dimension of the explanation for the link between perceptual exploration and reading is suggested by considering the process a child must use in learning and using decoding skills. Elkind (1975) has argued that the concept of units of reading, e.g., of letters, sounds, words, must be constructed before phonics can be used effectively and that these depend upon perceptual regulations. The child must recognise that a single letter, e.g., "t", can be represented in several different ways - upper or lower case and in various type forms and that "when used in a word one and the same letter can be associated with different sounds (t in to and in the) and one and the same sound can be represented by different letters (sun, son)". The child is thus faced by "problems analogous to those he meets in dealing with figure-ground reversals and schematization of part and whole" (Elkind, 1975, p.540).

Seriation may be seen in many aspects of reading. Piaget (1967) has argued that the unit concept develops as seriation and classification combine. He explains the emergence of the unit of number in terms of operational thought.

> A whole number is in effect a collection of equal units, a class whose subclasses are rendered equivalent by the suppression of their qualities. At the same time, it is an ordered series, a seriation of the relations of order. Its dual cardinal and ordinal nature thus results from a fusion of the logical systems of nesting and seriation, which explains why true number concepts appear at the same time as the qualitative operations. (p.53)

Attaining the concept of units of print seems to be, in essence, the same process as attaining the concept of a unit of number. If this is so, it would draw upon serial ordering capacity.

Seriation also seems to be called upon at other points of the reading task. Reversal of letters, e.g., reading "b" for "d", depends upon the child recognising that the order of the parts makes the critical difference (until learning to read the orientation of objects has not altered their identity). The order of letters within words is the means of distinguishing many frequently confused words such as "saw" and "was". Furthermore, although the sequence of words is usually no necessary cue to their individual recognition, it is a vital cue to the overall meaning, e.g., the "blind Venetian" and the "venetian blind", is he "running for a train", or "training for a run". In addition not only must the order of print (in space) be grasped but it must be matched with the order of speech (in time). Perhaps the most complex seriation of all is the multi-dimensional evaluation, one against the other, of all the many parts of a printed message to understand the mind of another person revealed there.

It may, therefore, be seen that a closer examination of the tasks involved in the operativity measures and in reading reveals that the link between the two is not as unlikely as it might at first appear but possesses considerable plausibility.

An empirical test of this suggestion was made in a longitudinal study of children's awareness of the units of print, their conceptual reasoning (seriation and perceptual organisation) and their reading during the first three years of school (Watson, 1984). Conceptual reasoning was found in path analysis to exert significant direct and indirect effects on awareness of print. This relationship was substantial despite allowance for the effects of oral language and reading ability. Thus, although oral language and reading level (as an index of print experience) each played a part in the emergence of the awareness of units of print, no explanation of the metalinguistic capacity of school beginners seems to be adequate without reference to the newly emerging conceptual reasoning competence. The study considerably strengthens the case for a cognitive developmental theory of learning to read.

However, because the operativity variables are part of Piagetian developmental theory it has been assumed that these cognitive influences on reading follow a developmental pattern. No study has examined the nature of this pattern in the first five years of school in relation to reading and, since Piagetian theory is currently subject to considerable challenge, it is important that these developmental assumptions be given close empirical scrutiny. In what sense are the cognitive variables developmental? Are they indices of a cognitive structure which shows stage related change? Furthermore, although there is a good basis reported here and elsewhere (Waller, 1977) for accepting the existence of a relationship between operativity and learning to read, the precise nature of that relationship is difficult to determine. An alternative explanation of the development of

conceptual reasoning gives major prominence to oral language. The nature of reading as a language processing task suggests that a strong link exists between speaking and understanding oral language, on the one hand and writing and reading printed language on the other (Donaldson, 1978; Goodman, 1970). The critical question is whether oral language capacity can explain the conceptual reasoning called on in learning to read. May oral language and operational competence be distinguished in their relationship to reading? Is operativity a prerequisite, a facilitator or an incidental correlate of reading ability? The difficulty in studying causality occurs largely because cognitive development and learning to read are not unitary activities but are attained as part of extended processes and one time of sampling the data tells nothing about those processes. Only on the basis of a close examination of the developmental pattern can a more precise delineation of its relationship to the child's emerging reading capability be determined.

The Purposes of the Study

Arising from these considerations there are two purposes for the present study: first, to assess the nature of the development of the predictor variable, operativity in contrast with oral language and of the criterion variable, reading, during the first five years of school, and second, to examine their interrelationship during this period, in order to test a cognitive developmental theory of learning to read.

Developmental Nature of the Constructs Studied

Human development refers to change over time in the characteristics of the subject. Is any variety of change developmental? The answer to this question produces sharp theoretical division. Most would say "No", and require certain specific conditions. Some, however, would tend to say "Yes". Something of the range and nature of those theoretical differences needs to be understood.

Earlier work on child development has been concerned with age-related change which has produced a linear model and a unidirectional conception of sequence. However, more recent writings have sought to encompass the greater complexity of human change in evidence and there has emerged a plurality of views which can be categorized along a continuum from strong to weak. (Baltes and Nesselroade, 1979). Strong conceptions require many criteria for change to be classified as development, such as irreversible sequence of stages, goal orientation, structure and qualitativeness, as specified, for example, in Piagetian theory. Weak conceptions of development, such as in social learning theory (Bandura, 1971), assume a cumulative model of development and require little more than that change be reliably identified. The differences seem to be partly a reflection of the complexity of human functioning and partly the result of the theoretical commitments of researchers which guide the questions they ask. For the present work it seems unwise to narrow the range of conceptions of development but rather to use theoretical predictions which seem suitable to the class of behaviour under study and then assess the nature of the development as empirically revealed in the light of the theory.

Operativity, chosen for this study to assess conceptual reasoning, usually is considered to be nearer the strong end of this continuum while oral language, the other aspect of the predictor set, seems likely to be located nearer the weak end of the continuum. Bayley's findings (Bayley, 1970) of greater stability for verbal than for reasoning tasks over an extensive longitudinal assessment of mental abilities gives general support for this. However, more detailed specification is required and is of particular interest in regard to the Piagetian variables. Are they indices of a cognitive structure which participates in stage related change or are they more parsimoniously described as following such change concepts as Brainerd's "measurement sequences" (Brainerd, 1978)? A measurement sequence is

determined by the structure of the test and occurs "whenever each item in the sequence consists of the immediately preceding item plus some new things" (Brainerd, 1978, p.176). Brainerd argues that an invariant sequence of development does not provide prima facie evidence for the existence of stage progression since measurement sequences or other non-organismic explanations may be found which can account for change.

Piaget's "structure d'ensemble", which proposes a psychological unity with interdependence of the sets of operations linked as logico-mathematical groupings, is very difficult to verify and has been regarded with varying degrees of scepticism by researchers. The forty one responses following Brainerd's challenge to the stage concept (Brainerd 1978), provide a lively example. A full consideration of this issue is beyond the scope of the present discussion and is unnecessary since several scholars (Brainerd, 1978; Feldman and Toulmin, 1975; Flavell, 1971; Larsen, 1977; Wohlwill, 1973) have argued it at some length from several different positions along the strong-weak continuum of developmental conceptions. Yet a review of the present position, including a consideration of some recent longitudinal evidence on the matter, is necessary and will be provided by the discussion of three questions in regard to cognitive change in middle childhood. First, is the change abrupt or gradual? Second, is the change qualitative as well as quantitative? Third, are the subsystems of the hypothesised structure interrelated and, if so, in what way?

In regard to the first question - abrupt or gradual change - there is ample evidence that success on a specific concrete operational tasks does not mushroom overnight but emerges by slow steps (Flavell, 1971) and, from individual to individual, follows different timetables (Almy, Chittenden and Miller, 1966, Almy, Dimitrovsky, Hardeman, Gordis, Chittenden and Elliott, 1970). Thus,

gradual rather than abrupt change in operational skills over the middle childhood years is found (Tomlinson-Keasey, et al, 1979). If change is gradual there seems less basis for considering it as constituting a stage. However, gradual is a relative term; change which is gradual over two or three years is rather different from that which is gradual over twenty years or a life time. If behaviour alters in its quality over two or three years, even if that change is gradual, there may be some basis for a stage conception (Ausubel, 1957). This leads into the second question which concerns qualitative change.

In an illuminating essay on the concept of continuity in Western thought, Kagan (1980) has suggested that the predominant assumption of smooth cumulative transitions and of long term change has prevented a recognition of the importance of shorter term change and hence of discontinuities in development. Psychology's linear, unidirectional growth models as applied to cognition are being challenged by evidence which suggests multidirectional patterns. Piaget's theory, for example, which suggests that there will be several qualitative changes during infancy and childhood has found echo in some recent longitudinal studies. McCall, Appelbaum and Hogarty (1973) reanalysed data from the Fels longitudinal study of IQ in children between two and a half years and seventeen years. Distinguishing between the general level of a person's IQ and his profile contour (developmental inflections in performance) they found major inflections in IQ performance at 6 and 10 years and perhaps at 14 with the change at 6 years being the most compelling. Although psychometric influences (repeated testing and change in the tests) seem to have had some effect the evidence remained clearly in support of a marked surge in development at about 6 years discontinuous with earlier change. A similar conclusion that there is a qualitative change in learning performance at about the same age has been suggested by White (1965). Earlier

discontinuities have been discovered by McCall, Eichorn and Hogarty (1977) in their reanalysis of the Berkeley Growth Study covering the first five years of life and by Uzgiris (1976) in a review of studies of cognitive change during infancy. Similarly, qualitative differences in early adolescence have been found in the study of concepts over the 9 to 13 year period (Nassefat, in Wohlwill, 1973).

There is therefore considerable evidence that qualitative as well as quantitative change does take place in the development of human cognition. In the move to concrete operational intelligence it is seen clearly in the difference between the perceptually based response of the typical four year old and the conceptually based response of the typical seven year old to a task such as the conservation of mass. This qualitative difference between responses seems to mark off different levels and thus gives some substance to the concept of stage. However, a number of unrelated qualitative changes would not be sufficient for the title "stage". They must have something in common which ties them together and which provides for economy and consistency of operation. This is assumed to be provided by some underlying mechanism that integrates the range of behaviours into a harmonious functional whole. This leads to the third and most difficult question about change - the interrelationship of the subsystems.

If concurrent emergence is important for the concept of the inter-dependence of concrete operational components (and Flavell (1971) thinks it is) then the evidence provides little support for this aspect of a stage. The subsystems of operativity seem to emerge rather asynchronously. (Flavell, 1971; Tomlinson-Keasey et al, 1979; Arlin, 1981). Flavell suggests that only in the final assembly of the stage is the structure integrated and that transitional periods are marked by the unstable organization of components.

Wohlwill (1973) believes that Flavell has "allowed himself to be misled by the pervasiveness of asynchrony" and has adopted a position where the development of "supposedly related concepts occurs in an essentially indeterminate fashion, with only gross correspondence of chronological periods remaining as a hopelessly watered down restatement of the synchrony principle" (p. 239). Flavell (1977) shows some ambivalence over this as evidenced by the comparison of his introductory assertion of the existence of a "complexly organized system of interacting (cognitive) components" (p.4) with his later challenge (p. 249) to an almost identical statement of Wohlwill. Nevertheless he believes that stages will not figure importantly in future scientific work on cognition.

Concurrence may well be an unreasonable expectation in the light of the gradualness of development, of measurement confounds such as task specific difficulty levels and the possibility of latency effects (Danner and Day, 1977). The heart of the issue is the interrelationship of the suggested operative components over the middle childhood years and in regard to this some recent evidence suggests that structures become more stable and intercorrelated during that period. Tomlinson-Keasey et al (1979) found that the seriation and numeration of six year olds predicted an increasing range of other operativity tasks measured in four waves of testing at intervals of 4 months, 7 months, 12 months and a further 12 months. These two early Piagetian abilities predicted the other operations more powerfully than did the Metropolitan Readiness Test. Although the authors' causal claims for some relationships on the basis of cross-lagged panel analysis must be regarded with scepticism because of the inadequacy of that method of analysis (Rogosa, 1979), the time precedence built into the design of the study and the relationships found (multiple correlations frequently exceeded .5) suggest

that a path analysis applied to the data would uncover important causal patterns. The results show a degree of empirical coherence for the operations tested and afford some validation for the construct of concrete operations not only in its final form, as Flavell suggests, but also at what may be nodal points along the way, as Wohlwill (1973) proposes.

Further longitudinal illumination of the relationship within operational subsystems is provided by a recent two year study of Arlin (1981). In Kindergarten and Grade 1 testing of nine operational tasks, Arlin (1981) confirmed the general order of difficulty found by Inhelder, Sinclair and Bovet (1974) but also reported intraindividual differences in the process of acquisition. The coefficients of reproductibility and scalability for a Guttman scalogram analysis of the difficulty order of the nine tasks failed to attain the level required to indicate a valid and truly unidimensional and cumulative scale (Nie et al, 1970). Arlin hypothesises that a linear relationship of the nine variables is much less appropriate than a two or three tier pattern across which integration and consolidation take place. Within a given tier, individual differences in the order of acquisition seem to be the rule rather than the exception. However, this suggestion must be accepted with some caution. Although the data are consistent with it, a much closer study e.g., measurement every 3 or 4 months and use of more finely graduated scales, is needed to confirm the suggestion.

In sum, concrete operations show gradual rather than abrupt change which nevertheless seems to reflect a qualitative difference in functioning from earlier mental activity. Although this change may not show a great degree of concurrence over two, three and four years there is evidence of coherence and interdependence of its subsystems. The evidence is far from strong enough to support the logico-mathematical stage proposed by Piaget but may be closer to

the less exacting family resemblance structure proposed by Feldman and Toulmin (1975) and perhaps a little stronger than the three ring circus (with some coherence but much that is unrelated) of Flavell (1977). Although the concept of stage in some forms may not fit the data very well, the alternative concept, sequence, seems to fit it less well, failing to account for the demonstrated interdependence of different but analogous concepts or skills. Certainly more longitudinal and finer grained data is needed before the stage concept is abandoned or can be given sufficient support for wider acceptance. At present it seems well to retain it (with an open mind) as a valuable heuristic - a position Geneva seems to accept (Inhelder, Sinclair and Bovet, 1974).

The first hypothesis is thus concerned with the developmental patterns of the predictor variables and states that operativity and oral language follow different developmental paths in the early school years. Operativity shows a degree of developmental surgence as befits stage related change while oral language shows steady continuous change.

Relationship of Operativity and Reading

There can be little doubt that a relationship exists between operativity and reading. In numerous studies discussed earlier significant and generally moderate correlations were found between various forms of operativity and a range of reading and reading readiness measures. However, the nature of that relationship requires detailed elucidation.

Two studies illustrate this research. Fifteen early readers (who read without formal instruction) were found (Howes, Polk and Goldstein, 1980) to be

significantly better in conservation tasks than were fifteen matched non-early readers. The correlation of their operativity (largely conservation) before formal reading instruction and reading after one year of instruction was .38. Thus, although operativity seems to be one of the factors in the successful reading of these subjects, the nature of the relationship is not clear. Is it a causal link and if so what is its direction?

Arlin (1981) tested 121 children at the end of their kindergarten year on nine operativity tasks including some from each of the three operative subgroups, conservation, classification and seriation, and compared their reading 12 months later. The correlations were low to moderate ranging from .16 to .35 but, in a regression process, the operativity measures affected reading significantly above the effect of a reading readiness test. Although double seriation was the strongest single predictor, tests from each of the three subgroups contributed to the best prediction of reading and it is concluded that all three subsystems must be represented for a satisfactory characterization of operative level. There are some limitations to the assessment of operativity in this study, however, which must qualify the results. Neither multiple seriation nor perceptual regulation were included among the operative measures. Both have been shown to be important predictors of reading (Elkind, 1975; Lunzer and Dolan, 1977; Watson, 1984) perhaps as strong as, or stronger than, those in the set used by Arlin. In addition, the use of one task in each measure and a narrow scoring range of 0,1 or 0,1,2, does not allow the gradualness of development nor the important distinction between trial-and-error and operational responses (outlined by Piaget for tasks such as seriation) nor any latency of response, to be reflected in the results. Despite the restriction to the predictive power of the measures which these factors are likely to produce, the existence of a relationship between operativity and reading is evident. However, again the nature of this

relationship is not clear. Though the time precedence of operativity in this and several other studies (Dimitrovsky and Almy, 1975; Howes Polk and Goldstein, 1980; Lunzer and Dolan, 1977) may suggest that operativity is a causal factor, data which allow the testing of rival hypotheses (e.g., language learning hypothesis) and the more sophisticated analysis of longitudinal patterns (e.g., using structural models) is required to elucidate possible causal relations. Is operativity causally linked with learning to read? If so, is it a unidirectional or reciprocal relationship? Does it function independently of oral language (commonly held to be critical for reading and intellectual development)? How does an understanding of the units of print (language awareness) which seems to be causally dependent on operativity (Watson, 1984), fit into this network of relationships? Is there a changing pattern in the relationship of operativity, oral language, units of print and reading during the first five years of school? If so, what is the nature of that changing pattern? These are the questions which this study will seek to answer. The central concern of the study is with the pattern of relationship between operativity, oral language and reading during the first five years of school. The second hypothesis states that operativity exerts a causal influence on reading which is partly distinct from the effect of oral language.

Bringing Theory and Method Together

The critical importance of a convergence of theory and methodology makes the longitudinal approach highly suitable for this study. No other method is able to capture the developmental dynamism of cognition and reading during the first few years of school. However, there are several major difficulties in longitudinal research, especially when the method is applied to cognitive developmental variables, that must be faced if it is to be used successfully.

Three issues will be considered; first, the need to resolve measurement problems raised by the nature of the developmental variables studied, second, the assessment of the construct validity of these variables by using a reinterpretation of Campbell and Fiske's multitrait- multimethod analysis and third, the application of stability values from that analysis to examine individual differences and qualitative change revealed in the measures.

Measuring Developmental Capacities

Longitudinal research, concerned as it is with the study of time-related constancy and change, encounters problems additional to those usually associated with measurement of less changeable capacities. This is especially the case for variables towards the strong end of the developmental continuum such as operativity which, as has been shown above, can be expected to manifest gradual and qualitative change during the 5 to 9 year age period being studied in this project. Difficulties occasioned by the gradual and the qualitative nature of the change will be considered in turn.

The gradualness of the development of concrete operational capacities is frequently overlooked in their measurement. They are commonly scored on a two point scale, operational, (those who can do the task) and non-operational, (those who cannot); or perhaps on a three point scale with transitional as an intermediate category, (e.g., Arlin, 1981; Howes, Polk and Goldstein, 1980). However, the gradualness of the emergence of these capacities means that there are likely to be individual differences that produce a range of intermediate levels of response. In seriation, for example, as well as those who can and those who cannot arrange the items in order, there are those who can only do so with the simplest materials such as four different sized stick figures. Others can order 10 rods each 1 cm different in size but fail when the rods are 4 mm different in size. Others correctly order small groups of the rods

but fail to relate all 10 rods. Thus a recognition of the gradualness of development means its measurement must also reflect the gradations of response shown. This implies the determination of a certain order of difficulty for responses and the assignment of score values accordingly. A two or three point scale will overlook a great deal of response variance which may reduce the predictive power of the measure.

Qualitative differences in responses stemming from qualitative change in development can lead to overestimation or underestimation of capacity. An overestimation of capacity can result from failing to distinguish between trial-and-error and operative modes of response to such tasks as seriation (Inhelder and Piaget, 1964). In an operative response to the single seriation task the child realizes the relationship of the rods before he begins to assemble them and so orders them without redundant comparisons. In a trial-and-error response the child does not seem to have this mental construction but physically compares and alters the elements in the emerging pattern to complete the array. If no distinction in the score is made between these two qualities of response an overestimation of the trial-and- error response will result. The problem of distinguishing between the two qualities of response can be minimised by a behavioral definition of the difference and by careful training of testers to discern it. Inter-tester agreement on this distinction was found to be 89% in regard to a multiple seriation task (Watson 1979).

An underestimation of capacity can result from failing to provide for response latency in the structure of the test. One aspect of developmental change revealed by research into Piagetian variables is that a response may be present in the subject before it is spontaneously available for use in a specific situation (Danner and Day, 1977; Stone and Day, 1978). Inhelder,

Sinclair and Bovet (1974) seem to be referring to the same issue when they comment that it is sometimes sufficient to ask the right question at the right moment in order to activate dormant mental processes. To allow for response latency the question may need to be repeated a number of times as Inhelder et al (1974) suggest or, after requiring a spontaneous response, the problem can be repeated in a more structured form so that greater support is given for the correct response. For example, in multiple seriation, the subject can be required, first, to place all pieces on the matrix (a spontaneous construction) and then, to replace several which the tester has omitted in a reconstruction of the solution, and then again to construct the whole matrix. If the subject fails in the spontaneous construction through response latency (a capacity present but not available) these next two steps are likely to increase the chance that the response will be elicited.

These difficulties in diagnosis, springing from individual differences induced by the gradualness and the qualitative nature of development, apply in any one time of testing. There is, however, a further difficulty when developmental capacities are compared from one time to another. Their fundamental nature may be changing during the period in which they are being monitored. If the change is largely unidimensional and quantitative in nature, established measurement techniques seem to be satisfactory to monitor that change over time. Vocabulary development measured by a test such as the Peabody Picture Vocabulary Test seems to be an example. However, when change is not unidimensional and is qualitative in its nature, measurement of that capacity presents a fundamental methodological problem for the researcher. How can such behaviour be measured to allow comparisons over age along a common dimension and yet at the same time reveal the qualitative change taking place (McCall, 1977)? If the nature of the test does not change it may not capture the qualitative behavioral change. If the test does change it may not

afford a meaningful comparison with results before the change. Some variables, while changing quantitatively over a lengthy period, also show shorter periods in which there is qualitative change. McCall et al, (1973, 1977) found this in an analysis of patterns of mental change in the data of the Fels Longitudinal Study and the Berkeley Growth Study. In the present study this seems to be likely for the test of perceptual regulation which is predicted to show gradual change from before the concrete operational period and beyond it, but, being partially isomorphic with the operational processes and strongly influenced by them, will undergo qualitative change in middle childhood (Piaget, 1969). If this is the case, a steady increase in mean scores can be predicted during middle childhood but a heightened level of individual differences shown by a lower stability is likely as concrete operations influence it.

Other variables such as are measured by operational tasks e.g., single seriation, seem to show little or no change up to about four or five years of age then undergo a period of rapid development over two or three years after which virtually all children can complete the task and there is no more change. However, as shown above, there seems to be greater gradualness than appears on the surface which is related in part to task difficulty factors such as whether the rods are 1 cm or 4 mm different in length. If researchers wish to measure subjects over a five year period on such a dimension they must vary the form of the test with age, drawing on theoretical considerations to ensure its appropriateness to the particular development of the subjects.

In sum, two different varieties of developmental change have been identified which require somewhat different measurement provisions and which are expected to show different developmental patterns in longitudinal investigation. First, there is development which is continuous, largely showing quantitative differences over time, e.g., vocabulary. Second, there

is development which shows qualitative differences during middle childhood, either together with continuous change, e.g., perceptual regulation or development which seems to reflect discrete qualitative differences in funtioning, e.g., single seriation. In the longitudinal measurement of all varieties of developmental change construct validity is a critical consideration but in the measurement of capacities which show qualitative change and which require a change of task in the measurement process, construct validity is even more important. If the form of the test is changing how can we know that what we are measuring over time is indeed the same thing? This calls for a consideration of how construct validity may be assessed in longitudinal research, an issue raised in the next section. This will be followed by a discussion of a related issue; the validity of the proposed variable groupings to be used in further analysis. After this the question of how these different patterns of development can be studied empirically will be discussed.

Construct Validity: Multitrait-Multitime Analysis

A useful technique for the study of validity is the multitrait-multimethod analysis (MTMM) proposed originally by Campbell and Fiske (1959) and discussed more recently by others such as Sullivan and Feldman (1979) and Marsh and Hocevar (1983). To assess the construct validity of longitudinal data, some reinterpretaton of multitrait- multimethod analysis is required. It has been used to study two aspects of validity - convergent validity and divergent validity - and to assess the measurement bias components of results. Convergent validity is indicated when different methods of measuring the same trait are substantially correlated. If different tests are measuring the same construct, a convergence of results can be expected. Divergent validity is indicated by a lack of correlation among theoretically separate constructs.

If similar tests are measuring different constructs, a divergence of results can be expected. To establish the matrix, at least two different measures are used of at least two different traits.

An assumption of the technique is that there are no time-induced differences in the subject's responses. Thus the tests are to be administered at the same time period. If, as in the present data, the assumption of no time effects upon responses does not hold, but instead the purpose is to study those time effects, the nature of the analysis is changed and the interpretation requires modification.

Campbell and Fiske (1959) suggest that four aspects of the multitrait multimethod matrix bear upon validity. First, the correlations between different measures of the same trait (validity diagonal values) should be statistically significant and of sufficient size to warrant further investigation. Second, each validity value should exceed the correlations between that trait and different traits assessed with different measuring methods. Third, each validity correlation should be greater than those between that trait and other traits assessed by the same measuring method. Fourth, the same pattern of trait correlations should be shown in all the method by trait matrix triangles.

The application of this technique for repeated measures of potentially unstable developmental characteristics requires some reinterpretation of these four criteria. For the data of this study, method difference is replaced by time difference; the same method is used to test each variable but with one year between each testing. The data thus produces a multitrait-multitime matrix (MTMT) for analysis. Without different measures of the one variable, methods bias cannot be assessed and the convergent correlations thus indicate

stability rather than validity. Assuming there is strong reliability, (internal consistency) substantial correlations in the trait by time diagonals become the basis for the first criterion which now indicates convergent stability.

The nature of this stability coefficient and how it is related to reliability and validity need careful consideration. Reliability and validity, as Campbell and Fiske (1959) point out, can be seen as regions on a continuum. In this context, reliability is agreement between two tests of the same trait using maximally similar methods. Validity is agreement between two tests of a trait using maximally dissimilar methods. Stability seems to occupy the middle ground partly shared by both and partly independent of each. Since the same test is used despite the time separation, the stability value represents a lower limit of reliability. Because of the time interval involved there are likely to be differences in the nature of the variable being tested. Has it changed beyond recognition or is it sufficiently similar at a later testing to be distinguished from other variables measured in the same time pattern? Thus, the stability value and its comparison with the divergent validity coefficients, give an index of construct validity. In practice even two different measures of a trait are not entirely independent and relative validity is always the best that can be attained (Campbell and Fiske, 1959).

Although the first criterion of Campbell and Fiske, when transposed into a multitrait-multitime analysis, is not strictly convergent validity the second and third criteria indicate divergent validity in MTMT. They are concerned with whether the measures are distinct from measures of other related variables over time. Since time-change does not seem to present as severe a test of convergent validity as method-change, the demonstration of construct validity is probably not as strong for MTMT as for MTMM.

The second reinterpreted criterion is that the stability value should exceed the correlations between that trait and different traits measured at different times. On the MTMT matrix this means that the stability diagonal value should be higher than the values lying in its adjacent column and row in the different-time blocks. In other words, the correlation across time (stability value) should be higher for the same traits than for different traits.

The third restated criterion from Campbell and Fiske is that the stability value should be greater than the intercorrelations of different traits measured at the same time. From the matrix a value in the stability diagonal is compared with those in the different-trait, same-time triangles. An assumption of this criterion is that traits are not highly correlated independent of method. This is difficult to assess with only one method.

The fourth criterion restated for MTMT is that the same pattern of correlations should be shown in each of the trait by time triangles. This assumes that, although the magnitude of the values might differ, their pattern reflects the strength of causal connections among the traits. However, with time changes it is possible that the causal links between variables will change. Indeed, it is a purpose of this study to examine these causal links and their possible changes. Thus the fourth criterion may not indicate construct validity. However, given an acceptable level of reliability and validity for the measures, the fourth criterion may shed light upon any changing causal relationships over time in the data. In studies such as the present one where variables are being grouped as predictors for further analysis, the relationships implied in this fourth criterion can be the basis for testing the validity of those groupings.

The implications of various result patterns when applying multitrait-

multimethod analysis to test construct validity in time sequence data require examination. If the stability coefficients are substantial and consistent and are distinct from the correlations of other variables (criteria 2 and 3) there is some support for the construct validity of the measure and an indication of stable development (given increases in means) during that period. This gives confidence for its use in further analysis. Consistent high stability of a test and increases in the mean shows that subjects are developing at about the same rate in that characteristic. If, on the other hand, the stability values for a variable are low and not well distinguished from other values, the instrument or the construct validity of the trait, or both, may be seriously in doubt. Alternatively, there may be qualitative change in the trait or ceiling effects of the measure. Further consideration of possible reasons is required.

In conclusion, the reinterpretation of multitrait-multimethod analysis for testing the construct validity of time sequence data changes its significance in important ways. Instead of construct validity values it provides stability coefficients which, although not giving as stringent a test as multitrait-multimethod analysis, can indicate construct validity. These stability values are an index of construct validity when compared with the values of other unrelated traits measured at the same time or at different times (Campbell and Fiske's first, second and third criteria).

The pattern of correlations in each of the trait by time triangles (Campbell and Fiske's fourth criterion) can allow a test of the validity of the variable groupings in the present study. This needs more detailed consideration which will be provided in the next section.

Validity of Variable Groupings

When two predictor variables are combined as one it is important but not sufficient to consider the construct validity of each separately. The question of the validity of the grouping must also be investigated. This can be achieved by using multitrait-multitime analysis to ask whether the variables, although distinct, are more strongly related to one another than they are to other variables measured at the same time or at different times. This is a stringent test since it is not assessing different methods of measuring the same variable nor even the same method applied at different times but rather is it testing different (though theoretically related) variables perhaps measured with different methods at the same time and at different times. Although no other example of this test has been found in the literature that might serve as a precedent, four criteria, based on the multitrait-multitime matrix and in part similar to those of Campbell and Fiske, seem appropriate.

The first is whether the correlation of the partners (i.e., the theoretically grouped variables) is significant and substantial at same-time and across-time measurement points. Does the strength of their convergence suggest that the validity merits further investigation?

The second criterion concerns the discriminant validity of the grouping as shown by the relationship of variables measured at the same time. Is the correlation of the partners stronger than the correlation of either with other variables measured at that same time? This involves a comparison of values within the same-time triangles adjacent to the matrix diagonal.

The third criterion concerns the discriminant validity as shown by the relationship of variables measured at different times. Is the correlation of a variable with its theoretically determined partner stronger than its

correlation with other variables measured at the same time as its partner i.e., in the same row and column of its matrix box?

The fourth criterion is concerned with whether the relationship of the variable partners across time is greater than the relationship of either of the variables with non-related variables measured at the same time. This means the correlations of partner variables in the different-time boxes will be greater than the values within the same row or column of corresponding same-time triangles. This is a very stringent criterion to apply to time-lag developmental data since the developmental process is likely to reduce across-time values.

Individual Differences and Qualitative Change

As well as being the basis for assessment of construct validity, the stability value can also have another important use in developmental time-sequence data. It can be an index of developmental change when compared with the other values it takes across time and with the stability pattern of theoretically contrasting variables. In this way, given sound reliability (in the sense of internal consistency), it can provide an assessment of the effects of individual differences in developmental change which may suggest the existence of qualitative as well as quantitative change in the data. It is this possibility which will be discussed in more detail now.

Two different forms of developmental change have been distinguished which seem to reflect greater and lesser degrees of continuity and qualitative change. Some behaviours seem to emerge with continuous quantitative increments throughout childhood, e.g., oral language. Others are likely to reveal the effects of a marked qualitative transformation in middle childhood. Of these, some undergo qualitative change while also showing

elements of continuity, e.g., perceptual regulation. Others, such as responses to operative tasks, are first shown in middle childhood, are then mastered by virtually all children during the next two or three years and, in that form, do not change again, i.e., they show discrete, qualitative change. The problem for the researcher is how these developmental continuities or discontinuities can be distinguished by empirical assessment. Can quantitative data reveal qualitative change? Must not any quantification of qualitative change gloss over and lose the impact of the changing organisational substructure? For example, it is often felt that to quantify Piagetian concepts inevitably forces them into an ill-fitting empiricist mould. If the purpose is to index individual intellectual development this is the case. Elkind (1981a) has cautioned against the assumption that Piagetian measures are comparable to psychometric assessments and rejects the translation of developmental assessments into psychometric values as a means of indexing mental growth. Developmental assessment records evolving organised wholes or "forms" while psychometric assessment deals with measurement units or "traits". Elkind's suggestion of a two part definition of general intelligence, "g", in terms of "gf", a form component and "gt", a trait component, seems to be a useful response to the call of Wohlwill (in Kagan, 1980) for a major reorientation in the assessment of cognitive functioning to replace inadequate linear models.

However, if the purpose is not individual assessment but the developmental study of change, McCall (1977) has described how qualitative change can be detected in quantitative data. This can be done by distinguishing between developmental function and stability of individual differences in longitudinal data and McCall has illustrated the usefulness of this distinction in his reanalysis of major growth studies (e.g., McCall, Eichorn and Hogarty, 1977). Developmental function refers to the increase or decrease in frequency of a behaviour across age and may be continuous or discontinuous. The stability of behaviour refers to the degree to which the rank order of individual subjects

on a measure corresponds to their rank order on that same measure (or a conceptually equivalent one) assessed at a later time. The correlation of scores across age is presumably an index of individual differences in developmental progression for that capacity. A high correlation indicates that there is little change in the rank order of subjects over the period. A low correlation (with satisfactory test reliability) indicates greater change in rank order and suggests that individual differences in the development of that capacity are considerable. For that capacity it may be a time of very marked developmental ferment and indicate a qualitative change in functioning. As the capacity becomes settled in its development the stability values are predicted to increase (given substantial variance and reliability). Since correlations are assumed to be independent of the means of the distributions from which they are calculated, developmental function and stability are potentially independent of one another and each adds importantly to the understanding of developmental change. The common longitudinal strategies of examining the stability of variables while ignoring changes in the average performance over ages, means that such studies are not truly developmental in nature (McCall, 1977). The use of stability as a potential index of qualitative change does not reduce the need for description of the logical substructure of responses. On the contrary, findings obtained in this analysis may underline the importance of providing a "form" component in the representation of intelligence.

It is in the assessment of changing relationships over time that longitudinal design has overwhelming advantages over cross-sectional research and multitrait-multitime analysis is a valuable technique to study this. It might be wondered, finally, how the same stability value can be used to indicate validity at one moment and developmental change at the next. The difference lies in the comparisons being made with it. Discriminant validity of a trait is indicated by whether its prediction of itself is distinct from

its prediction of other traits. The nature of developmental change is indicated by whether its prediction of itself is consistent (given increments in the mean) and by comparison with the levels of stability of other traits.

Hypotheses of the Study

Finally, before considering the details of research procedure, the hypotheses need to be stated in terms of the principles of assessment discussed in this section. The first has to do with the differing nature of the development of the predictor variables and the second with their distinct relationships to the criterion.

1. Operativity and oral language, although both increasing rapidly in the early school years, develop in different ways. Operativity shows considerable individual differences in its emergence consistent with developmentally surgent, stage related change. Oral language shows a lesser level of individual differences indicating developmentally steady, continuous change.

2. Operativity exerts a causal influence on reading which is partially distinct from the effect of oral language and from awareness of units of print.

Method

Design

The longitudinal design of the study involved five waves of testing at the end of each school year from Kindergarten to Grade 4 as shown in Table 1.

PLACE TABLE 1 ABOUT HERE

Sample

The original subjects were drawn by chance procedures from three adjacent public schools in a middle-class southern area of the city of Sydney. In two schools pupils had been allocated to non-graded classes by chance from enrolment lists. In one of these schools two of four kindergarten classes (65 out of 101 children) were selected and in the other school, because there were fewer facilities for testing, one of five classes (33 out of 164 children) was selected, in each case after ascertaining the willingness of the teacher to participate in the project. At the third school pupils had been apportioned to classes on the basis of preliminary skill tests and a different procedure was followed. The 50 children were drawn from all three classes by selecting every second pupil on alphabetical class lists until the sample was complete.

A major problem of longitudinal research is sample wastage during the extended assessment period involved. This was kept to a minimum by following subjects who moved to areas within reasonable proximity to the original schools. At the five annual testing periods the sample size was 148, 142, 136, 135 and 128 respectively. In no one year were more than 6 pupils tested away from their original school. In the first year of testing there were 69

boys and 79 girls with a mean age of 5 years 10 months while in the last year there were 54 boys and 74 girls with a mean age of 9 years 10 months. Subjects were lost to the sample largely because of a change in parental work location. The wastage rate over five years is notably low and does not seem to be a significant factor influencing results.

The repeated testing of the same sample over five years introduces the possibility that the testing experience itself will have a significant effect on subsequent performance on the tests and thus be an uncontrolled influence on the relationships being studied (Wohlwill, 1973). For example, retest effects of intelligence measures were found to be considerable in a large adolescent sample (Nesselroade and Baltes, 1974). It is often thought that, when testing waves are separated by one year as in the present project, this effect is not likely to be significant. However, in the study of Nesselroade and Baltes (1974) there were significant effects despite a 12 month interval between testing and thus it is desirable to monitor this influence. An ideal technique is to administer all tests to a fresh matched control group at each wave after the first. This, however, presented a demand beyond the resources of the project and so a compromise was struck by enrolling 44 new subjects in the third year of testing (22 girls and 22 boys) from the same socio-economic group who attended a public school 3 km from the nearest core sample school. Average age of the control sample was 8 years by comparison with the core sample average of 7 years 10 months.

Materials

Reading measures. Two criteria were of particular importance in guiding the choice of reading tests from amongst the many available: suitability for an Australian sample and discriminability across the five year age change of the subjects during the project. An Australian revision of the Schonell

Graded Word Reading Test, the St. Lucia Graded Word Reading Test (Andrews, 1969) was chosen to assess word recognition. The test is a well established individual measure consisting of 100 finely graded single content words. The subject orally reads the words until the cut off point, 10 incorrect responses, is attained. The test has shown good discrimination from grades one to six and consistent high reliability.

Reading comprehension was measured by a test developed for use with children aged six to ten years in New South Wales Schools, the Paragraph Understanding Test (Hall and Pacey, 1978). It is a 20 minute group measure consisting of 33 graded paragraphs each with a multiple choice question. To eliminate the effects of guessing, a cut-off point was applied after subjects scored 8 out of 10 responses wrong. In the fourth and fifth years of the project an Australian developed cloze test, GAP Reading (McLeod, 1977), was added to provide a wider spectrum of reading measures and allow further assessment of the convergent and divergent validity of the initial measures. In addition, at the fifth year, the Reading Comprehension Test of the Progressive Achievement Tests (Elley and Reid, 1969), a New Zealand test adapted for Australian use, was also administered to all subjects as an additional means of monitoring construct validity.

Piagetian measures. Two indices of operative thinking, Seriation and Perceptual Regulation, were used in this study. Seriation consisted of a three part test; part one, Single Seriation (Lunzer, 1970), part two, Multiple Seriation with 9 pieces on a 3 x 3 matrix board and part three, Multiple Seriation with 25 pieces on a 5 x 5 matrix board (Watson, 1979). The first and second parts were administered in all five waves of testing with one exception. This occurred at the fifth testing occasion when Single Seriation, for which the subjects, mostly then over nine years of age, had attained a ceiling, was omitted. The third part, Multiple Seriation 5 x 5, was found to

be too hard and impractical to administer at the first and second waves and was administered in waves three, four and five as an extension of the task difficulty of the measure. The appropriateness of this change in task to match the expected change in capacity was suggested first from the finding (Lunzer, 1970) that a similar seriation task was not attained by more than 50% of subjects until age 9 and was confirmed in trials of the materials.

The 25 piece 5 x 5 matrix multiple seriation task, which followed the 9 piece task, presented the child with 25 fish graded from bigger to smaller on one dimension and from darker to lighter on the other. The pieces were mixed and placed on a table before the child next to the empty 5 x 5 matrix board. The subject was first required to construct the matrix spontaneously ("put the pieces on the board where they belong"). To overcome the problem of under-estimation of capacity due to response latency, the instruction was given as many times as was needed, a prompt was provided if the response was incorrect and the task was repeated in a graded series of easier forms in the next step of testing. In this the tester constructed the whole matrix correctly without comment then asked the subject to replace 4, 8 and 12 pieces which the tester removed in turn. Next, all the pieces were removed and the child was asked to reproduce the whole matrix. Finally, the subject was shown the board with all but the top row and the left hand column of fish removed and was required simultaneously to co-ordinate the two dimensions of the task by placing three non-adjacent fish correctly on the board. As well as providing for response latency, this graduated presentation, with a maximum score of 15, sought to allow for the range of intermediate levels of response capacity which, as shown in earlier discussion, is likely to exist in a sample. The problem of over-estimation through failing to distinguish between trial-and-error and operative responses was attacked by providing higher scores at three parts of the test for operative rather than trial-and-error constructions.

Perceptual Regulation, which was assessed by the Perceptual Ambiguity Test developed by Elkind and Scott (1962), provides an indication of how well the child can go beyond the more salient aspects of a complex visual field and perceive other elements of the stimulus. Seven ambiguous line drawings (e.g., two face profiles which, when juxtaposed, make a vase) are presented one at a time to the subject who is asked what he can see. If there is no response the question is asked again until the child replies or says he cannot see anything. If one aspect of the drawing is reported e.g., two faces, the child is asked can he see anything else. The question is repeated if there is no answer. If the child says "no" the next drawing is presented. One point is scored for each item identified with a maximum of 21 for the seven items.

Language Measures. To assess oral language, a listening test requiring an oral response (Grammatic Closure, Kirk, McCarthy and Kirk, 1968) and a vocabulary test (Dunn, 1959) were administered and scored following the standard procedures. Language awareness was assessed in waves two and three of the testing with the Units of Print Test developed for an earlier study (Watson 1984). Subjects were found to have attained a ceiling on this test after grade two (wave 3) and it was thus inappropriate for administration at waves 4 and 5.

Procedures

Each wave of testing was conducted in two phases, individual tests and group tests. The individual measures were administered mostly in the school library by a team of some ten undergraduate research assistants who were trained by demonstration and supervised rehearsal until a criterion level of competence was achieved (usually after 4 to 6 administrations). To provide greater standardization of procedure each assistant or group of assistants administered one measure and subjects were moved from test to test. This also ensured that no one testing period was excessively long, that there was a

change of posture, tester and activity every few minutes and that results of one test did not contaminate scores on another. The average time of the individual tests ranged from 3 to 8 minutes and the total average time was 25 minutes. In some cases where assistants administered the criterion measure, St. Lucia Word Reading, and also some predictor tests, steps were taken to prevent any influence on results by ensuring that no tester administered both measures to the same child. The use of specialised testers and the variable time taken for the different tests, however, meant that it was not possible to ensure tests were administered in a predetermined order. As a subject completed a test he or she was directed to another tester who was free. The order of administration, thus varied by chance, could influence the results but it seems most unlikely given the minimum generalisation found to take place across cognitive tasks (Brainerd and Allen, 1971) and the lack of any significant effect found (Hooper, Toniolo and Sipple, 1978) when six different presentation orders for Piagetian tasks were compared. The research assistant team did not remain the same across the five years although the majority continued for three or more years and some for all five years. At the beginning of each testing period, however, there was a time of procedure revision and practice for continuing testers and training for new team members. At all times the researcher worked closely with the team directing and supervising procedures. Steps were taken to ensure that the children knew about the nature of the testing and were relaxed and free of distractions as they responded.

The group measures were administered in a classroom by the researcher and an assistant to some 25 subjects at a time seated in rows of desks with screens positioned to prevent any subject seeing the work of others.

Data Analysis

The pervasiveness of measurement error and its potential influence on the interpretation of data (Goldstein, 1979) means that the first analysis needed is one of the reliability of all measures.

Since reliability is specific to the sample and to the particular testing occasion, measurement error variance defined as internal consistency was calculated using Cronbach's coefficient alpha on the basis of the full sample for each value (Goldstein, 1979) using an SPSS programme (Hull and Nie, 1981). Two of the measures with variable ceilings depending on student responses, the Peabody Picture Vocabulary Test and the St. Lucia Graded Word Reading Test, presented problems for the calculation of alpha since it requires equal item totals for each subject. To overcome this problem three coefficients were calculated for both tests each using a different cut-off point. The first point was that to which every child had completed, the second point was that to which 50% had completed and the third point included all responses of the highest scoring subjects. There was not a great deal of difference between the coefficients calculated in this way and the second cut-off point, which usually produced the lowest coefficient, was used as the most satisfactory basis for calculating reliability in these cases.

Calculation of the means and standard deviations, the t-test of the difference between the core sample and the control sample, the simple correlations for multitrait-multitime analysis and for predictor-criterion relationship and the multiple regression analyses were completed using SPSS computer programmes (Hull and Nie, 1981; Nie et al, 1975).

To allow for the graphical comparison of change over the five year period all means for each measure were reduced to standard scores with a mean of 0

and a standard deviation of 1 before being plotted on graphs. A multitrait-multitime analysis was provided by arranging the simple correlations in a matrix form as suggested by Campbell and Fiske (1959, see earlier discussion) both for all criterion measures and for the reading tests. This allowed an examination of both construct validity and developmental stability.

The relationship between the predictor variable groupings and the criterion variables was studied first using regression coefficients in a backward regression procedure (Hull and Nie, 1981). All variables were first entered into the equation then the different variable groups, in turn, were omitted. The significance of the drop in the multiple correlation was calculated and used to assess the contribution of that variable group to the regression.

Path analysis was performed to represent and summarise the overall relationship of the predictor and criterion variables and to attach quantitative estimates to causal effects postulated on the basis of the theoretical model. The firm theoretical foundation for the present study, its longitudinal design and the usefulness of determining causal patterns in the relationships makes this form of analysis appropriate (Duncan, 1975; Kerlinger and Pedhazur, 1973; Wolfle, 1980). To handle the complexity of the inter-relationships of the multiple variables over a five year period, the raw scores for each year were standardised and the composite measures of operativity, language and reading were computed as the basis for the structural analysis.

Path coefficients were calculated using COSAN (McDonald, 1980), a program developed to analyse covariant structures using maximum likelihood estimation procedures. The goodness of fit of a model was assessed by two methods, the overall chi-square statistic and the distribution of residual correlations.

χ^2 measures the difference between the theoretical model and the observed data. The larger the value of χ^2, the greater the difference. As a test of the signif^icance of the difference, a p value is given. When, for example, $p < .05$, the difference between the data and the model is significant, i.e. the data do not fit the theoretical structure. When $p > .05$ the difference between the data and the model is insignificant and the structure is accepted. The COSAN program which is able to provide path coefficients for the whole model in one computation and gives a residual coefficient matrix for the estimation of fit, is more efficient with a large number of variables than the use of ordinary least squares regression procedures.

Results

The major focus of the data, as indicated, is on the relationship of operativity and reading during the first five years at school while controlling for the influence of oral language. In addition, the effects of the child's developing awareness of print on these relationships will be examined. Before this is done, however, consideration will be given to the possible effects of repeated measurement and historical time period on the results. It is then proposed to explore the variables' construct validities, and the validities of the proposed variable groups. For this, multitrait-multitime analysis will be used. The nature of developmental change will next be studied, principally by an examination of increments in the means and of the pattern of stability values over the five years. This will lead to some reconsideration of construct validity of the predictor variables. Finally, the relationship of the predictor and criterion variables will be considered

using simple correlations, multiple regression and path analysis. The strength of the conclusions is considerably increased by the multiple comparisons provided by the longitudinal nature of the study which allow general trends in the data, rather than individual salients, to be the basis of such conclusions.

Effects of Repeated Testing

To assess any effects on the core sample of repeated experiences of the same tests, 44 additional subjects, a control sample matched with the original children, were given the full battery of tests in the third wave (Table 2). For neither the reading tests nor the Piagetian measures, (Seriation and Perceptual Regulation) were any significant differences between the groups shown. Of the language measures, Units of Print and Grammatic Closure showed no significant differences but in the Peabody

PLACE TABLE 2 ABOUT HERE

Picture Vocabulary Test, the control sample, with a mean of 74.95, scored significantly better than the core sample which had a mean of 71.31 ($p < .05$). This difference, in favour of the 44 new subjects, is unlikely to be a depressor effect of testing but perhaps registers a difference in school vocabulary teaching between the two groups. These results, however, indicate that three yearly waves of repeated testing gave no special advantage to the core subjects over those who were new to the tests. It seems that the school program, which involves activities and materials similar to the testing procedures, effectively prepared the children for assessment of this kind and that the additional annual testing of the core subjects made little appreciable difference at least at this third testing wave.

This finding is in accord with results obtained by Hooper, Toniolo and Sipple (1978) who checked repeated testing effects on operativity measures. They enrolled 48 additional subjects (original sample 102) in the second year of testing and found there were no significant differences in the performance of conservation and transitive inference tasks across the two groups. Although resource limitations in the present project have meant that retest effects could not be monitored over all years, this check gives some basis for the assumption that repeated measurement is not a significant influence on the results of the present data.

Effects of Historic Time Period

Another potential problem related to sampling is the differential effect of historical time period on development. There is some evidence (Nesselroade and Baltes, 1979) that different birth cohorts may undergo different experiences that produce significant effects on intellectual development. No attempt was made in this present study to examine possible cohort effects by including other school year cohorts beyond the one selected and this may be a limit to the generality of the findings. However, McCall (1977) feels that this concern is overstated since neither the inter-relationship of variables nor the stability of individual differences have been shown to be effected. The speculation of Baltes, Reese and Lipsitt (1980) that historic period influences operate most strongly in adolescence and early adulthood while normative age-graded influences are dominant in childhood seems feasible and suggests that historic time period effects on the present data may not be great. As McCall, Eichorn and Hogarty (1977) point out, age, year of birth and time of measurement are inevitably confounded and no methodology exists to separate them without making some assumptions. That there is little peculiar effect of birth cohort in this sample seems a reasonable assumption for the interpretation of the data. However, should the

results be compared with a cohort which experienced the direct effect of events such as a major war, an economic depression or a change of school admission age, the conclusion may need some qualification.

Establishing Construct Validity

The reinterpretation of Campbell and Fiske's multitrait- multimethod analysis for longitudinal data provides four criteria which will be applied to the predictor and then to the criterion variables, in turn, to assess construct validity. It can also be extended to yield four other criteria which will be applied to assess the validity of the variable groupings.

PLACE TABLE 3 ABOUT HERE

Validity of Predictor Variables. The first criterion of multitrait-multitime analysis is that the stability values (underlined in Table 3) be significant and of sufficient magnitude to warrant further examination of validity. Although in some cases there is considerable fluctuation of the stability of the same predictor variable over five years, all 40 coefficients are significant and every variable has at least three coefficients larger than .45. The language variables, Vocabulary and Grammatic Closure, are the most stable with correlations ranging from .40 to .75 and each has an average value of .58. The Piagetian variables, Perceptual Regulation and Seriation, are less stable with values ranging from .22 to .66 and each with an average value of .38. These lower levels of stability are not surprising given that this age period is a time of major developmental change according to Piaget's theory. For all variables there is reason to proceed further with the investigation.

The second reinterpreted criterion is that the stability value should

exceed the correlations between that trait and different traits measured at different times, i.e., in the adjacent column and row of the same block. Vocabulary, Grammatic Closure and Seriation all pass this test upon every comparison (60 for each). Perceptual Regulation meets the criterion on 55 out of 60 comparisons. The exceptions all occur when the year one measure predicts the responses at years three, four and five. Thus there is complete support for the discriminant validity of Vocabulary, Grammatic Closure and Seriation on this criterion. Support for the discriminant validity of Perceptual Regulation is less strong but is nevertheless quite firm.

The third reinterpreted criterion requires that the stability value exceed the correlations of different traits measured at the same time, i.e., coefficients situated in the adjacent value triangles. Vocabulary and Grammatic Closure pass this test on all 60 comparisons showing very strong divergent validity. Seriation passes this test on all but 3 of the 60 occasions and seems to possess good validity on this criterion. Perceptual Regulation passes the test on 54 of the 60 comparisons which seems to indicate a satisfactory level of discriminant validity.

The fourth criterion reinterpreted for multitrait-multitime analysis, the pattern of the relationships of the values, is relevant to the analysis of the variable groupings and to developmental change so is left for the discussion of those issues.

In summary, the measures of language, Vocabulary and Grammatic Closure, show a high level of stability and of divergent validity. Of the operative measures, Seriation is also stable although not as stable as the first two and possesses good discriminant validity. For the other operative measure, Perceptual Regulation, the stability, although significant throughout, fluctuates from low at first to high at later values and the discriminant

validity is by and large good. Overall, the measures may be seen to possess moderate to strong convergent stability and at least satisfactory divergent validity. Considered together these two attributes provide firm evidence for the construct validity of each of the four predictor variables. Reasons for the stronger construct validity of the oral language variables will be discussed in the section on developmental change.

Validity of Criterion Variables. To provide additional information on the construct validity of the main criterion variables, Word Recognition (St. Lucia) and Paragraph Understanding, a cloze reading test (GAP) was added to the battery in grades three and four and another reading comprehension test (Progressive Achievement Test) was added in grade four. The multitrait-multitime matrix (Table 4) reveals a strong and consistent stability for Word Recognition with six values ranging from .76 to .86 and an average value of .82. Its correlation with Paragraph Understanding and the other reading tests is moderate to strong, with values ranging from .38 to .67, yet it is clearly distinct from them with its stability value greater than its correlation with any of them at every wave of measurement.

Paragraph Understanding shows moderate to strong stability which tends to increase over time and which has values ranging from .40 to .64 with an average of .54. Although its convergent stability is not as strong as that of Word Recognition on this first multitrait- multitime criterion, it seems quite sound. Its divergent validity, in contrasts with other reading tests, however, is not strongly supported by the analysis.

In regard to the second criterion, involving discriminant validity, Paragraph Understanding's stability value is greater than its comparison values on 13 of the 21 occasions and on the third criterion is greater on only 4 of the 21 comparisons. Its discriminant validity does increase over time

when related to Word Recognition, however overall, it is not very distinct from Word Recognition and is less distinct from other reading measures. It seems that each is measuring a very similar capacity. Perhaps the distinctiveness of Word Recognition is to some degree a methods difference since it is an individually administered oral response test while the other three are group administered written response measures.

It is of interest to note that when comparisons of Paragraph Understanding with Word Recognition are one year removed not only is the convergent stability much stronger but the divergent validity comparisons are much more favourable. For the second criterion it passes all six tests and for the third it passes four out of six. When the comparison is two or more years removed on the second criterion it fails in five of the six comparisons and on the third it fails all six comparisons. There thus seems to be a puzzling change in the rank order of subjects in reading comprehension which does not show itself in word recognition. Perhaps this is a developmental characteristic which is related to reading comprehension in a way it is not related to word recognition. This issue however, requires elucidation on the basis of further evidence.

Thus, in regard to the criterion variables, the analysis shows that Word Recognition possesses strong convergent and divergent validity and Paragraph Understanding possesses good convergent validity but at best fair divergent validity. In addition the analysis does support the validity of combining the two main tests even though it suggests that not a great deal more variance may be ~~controlled~~ explained by the two than by either one. The validity of the predictor variable grouping will now be examined.

Validity of Variable Groupings. The previous multitrait-multitime analysis has shown that each predictor variable is distinct from all other

variables including its theoretically determined partner. It is now necessary to examine whether this theoretical grouping receives empirical support. Do they belong together? The oral language partners, Vocabulary and Grammatic Prediction, will be considered first, then the operative grouping, Perceptual Regulation and Seriation will be discussed. The criteria as adapted from Campbell and Fiske (1959) and discussed earlier in the section, "Bringing method and Theory Together, Validity of Variable Grouping", will be applied.

The first criterion applied to the language pairing requires their correlation to be significant and involves the 25 coefficients of Vocabulary and Grammatic Prediction shown in Table 3. The values range from .14 to .54 with an average value of .35 and all except one are significant at the .05 level. Oral language thus passes the first variable grouping test. The second criterion concerns the discriminant validity of the grouping and requires their correlation to be greater than that of either when correlated with other variables measured at the same time. This involves 20 comparisons in the same-time triangles of values adjacent to the diagonal in Table 4.3. Oral language passes on 19 of the 20 comparisons and thus may be deemed to possess very good validity on this score. The third criterion, another discriminant validity test, requires the correlation across time of a variable with its partner to be greater than its correlation with other variables measured at the same time as its partner. Of the 80 comparisons involved for oral language, it passes this test on 73 occasions demonstrating good discriminant validity. In regard to the fourth criterion, which compares across-time values of the variable partners and their same-time correlation with other variables, there are 80 comparisons. The oral language partners pass this test on 73 occasions, which, considering the stringent nature of this criterion, is an indication of very good divergent validity for that grouping. In sum, the oral language variables, Vocabulary and Grammatic Prediction, in their relationship to one another, are distinct though

significantly related and in their relationship to other variables, possess very good divergent validity. Thus, their grouping as measures of oral language on the basis of theoretical considerations receives very good empirical support.

The operative variables, Perceptual Regulation and Seriation, receive some empirical support as a predictor pair but not as strong support for their grouping as do the oral language variables. On the first criterion, which assesses whether they are significantly and substantially related, the range of the values is from .01 to .34 with a mean of .20. Of these 25 values 21 are significant at the .05 level. The relationship, although generally significant is not very substantial. However, a closer inspection of the matrix reveals that the four non-significant values occur in correlations with Perceptual Regulation measured in the first two waves of testing i.e., in Kindergarten and Grade 1. If the matrix is considered in two parts, with correlations involving Kindergarten and Grade 1 (waves 1 and 2) and those involving grades 2 to 4 (waves 3, 4 and 5) there is a striking difference. Correlations with Kindergarten and Grade 1 tests have a mean of .16 while those involving Grades 2 to 4 only, have a mean of .25. This, however, may not be considered a valid comparison as the correlations with the first and second waves include variables tested three and four years later while those with the third and fourth waves are correlated only with variables tested one or two years later. Thus the same test was applied with time-lag kept equal for both halves of the data, e.g., across no more than a two year span. This meant that the lower four blocks on the left of the matrix were excluded for this analysis. The results change very little with the mean correlation of waves one and two being .17 and of three and four, .25. Thus, although the correlations with the first two waves of testing of the operative variable grouping are not all significant and overall not very substantial, this pattern changes in the last three waves with all values significant and more substantial.

In regard to the second criterion, discriminant validity in relation to other variables measured at the same time (within the value triangles adjacent to the diagonal), the correlations of the two operative tests exceed their correlations with other variables in only 12 of the 20 comparisons. However, when the matrix is considered in two parts again, most of the failures in this criterion occur in correlations with the first two waves (5 out of 8 comparisons) while in the last three waves of testing there is a far smaller proportion (3 out of 12). A similar pattern emerges in application of the third criterion which assesses discriminant validity in relation to across-time correlation comparisons. Overall, of 80 comparisons the two operative variables fail to pass the test on 36 occasions. Of these, 34 occur in correlations with wave one or two measures out of 56 comparisons. Of the 24 value comparisons with correlations in waves three and four, only 2 fail this test. Again, holding time-lag constant for both halves, the failure pattern is still greatly disproportionate with 11 out of 24 in the first half and 2 out of 24 comparisons in the second. In regard to discriminant validity as assessed by criteria two and three, the overall result is poor. However, closer analysis of the data reveals that the lack of discriminant validity is found almost entirely in relation to the measures taken in Kindergarten and Grade 1.

The fourth criterion requires across-time correlations of the partners to exceed their same-time correlations with other variables. Values for Perceptual Regulation and Seriation fail this test in 31 of 80 comparisons overall. This breaks down, on considering the two halves of the matrix (columns 1 and 2 in contrast with columns 3 and 4), as 29 out of 56 for waves one and two and 2 out of 24 for waves three and four. Holding time constant as above, this becomes 11 out of 24 for the first half and 2 out of 24 for the second. Again, although the overall data comparisons are not good on this criterion, the ~~lower~~ poorer discriminant validity is found almost entirely in the

first two waves of testing. The two operative variables show good discriminant validity in waves three and four.

In sum, the validity of the operative grouping varies across time. In the first two waves, although most of the convergent validity values are significant they are not very substantial while the divergent validity is poor. At later waves all convergent validity values are significant and more substantial while the divergent validity is good. Bearing in mind the stringent nature of the criteria as discussed earlier, this does not disqualify the operative variables as a grouping, especially since the great majority of the values that throw doubt upon the validity of the grouping occur in predictions spanning three or more years. Nevertheless, it calls for an explanation of the difference between the lower validity with subjects six and seven years old and the higher validity at ages eight and nine. Four possibilities need consideration; variations in reliability, changes in task content, practice effects of repeated measurements, and the influence of developmental change.

A first consideration, that the difference is due to reliability, will bear some explanatory weight but not all. The alpha value for Perceptual Regulation is lowest in wave one (.64) but this is not so for Seriation which has its lowest value in wave five (.59). However, the strongest reliability values are shown in waves three and four which may be part of the reason for stronger validity in relation to those waves. Yet differences in reliability values are not very great and other explanations are needed.

Another possible explanation is that the difference is due to the variations in the Seriation Test content to match changing subject capacity (a 5 x 5 matrix task was added from wave three and single seriation was omitted at wave five). This certainly seems to be the reason why the reliability for

this measure is highest in waves three and four where it has the most items. However, the greater part of the lower discriminant validity is in relation to Perceptual Regulation which does not change its form at all over the five waves of testing. Thus, although change in the form of the Seriation Test may be a factor, it does not seem to account, in itself or in combination with the first explanation, for the striking difference between the earlier and later validity values for operativity. The difference is too great.

A further suggestion is that practice effects become evident in regard to operativity after two waves of testing and thus the results become more stable and distinct. However, the failure to find significant differences for operativity between the core subjects and the fresh subjects recruited in wave three, argues against this, as does the absence of this effect in the language variables. Both operativity and oral language tasks seem equally familiar to school pupils. Nor do the measurement procedures admit of any irregularities which could be an explanation of this difference.

All these suggestions even in combination do not possess sufficient explanatory strength to bear the full weight of the data. There seems to be another factor (or other factors) operating.

Perhaps the most important explanation is one that is consistent with the developmental theory proposed: In middle childhood, there is such major cognitive developmental change that it will inevitably produce fluctuating values such as these. This possibility will be considered in relation to a broader examination of the data in the next section.

Developmental Change

There are two principal sources of information on developmental change in longitudinal data - comparison of means and comparison of stability values - which, when considered concurrently, give a far more sensitive index of the nature of change than either on its own. The stability coefficients can provide an assessment of the effects of individual differences in developmental change and perhaps suggest the existence of qualitative change, as discussed earlier. This is very important for the present data in which two different forms of developmental change in the predictor variables, reflecting greater and lesser degrees of continuity and qualitative difference, are expected on theoretical grounds. The forms of change predicted are firstly, those revealing continuous quantitative increments (e.g., oral language), and secondly, those revealing marked qualitative disturbance, either together with continuous quantitative increments, e.g., perceptual regulation or as discrete qualitative change, e.g., seriation. In this study the means and stability values can be compared from one kind of variable to the next to determine whether this theoretically predicted pattern can be detected. This gives further valuable insight into construct validity because it considers whether the variables perform as they are theoretically expected to do. It may thus suggest a broadening of the idea of construct validity as developed on the basis of psychometric assumptions. In this section an additional assessment of the nature of developmental change evidenced in the MTMT analysis will be given by path analysis of the language grouping, the operativity variable grouping and the grouping of the reading measures. The predictor and the criterion variables will be considered separately.

Development of Predictor Variables The means and standard deviations for all predictor variables are shown in Table 5. Seriation, the content of which was modified to allow for changes in development, is shown in two

PLACE TABLE 5 ABOUT HERE

sections, single and multiple, 3 x 3, across all waves and multiple, 3 x 3 and 5 x 5, measured from wave three onwards. The fifth wave value for the first section of Seriation is, in part, an estimate based on the value of single seriation at wave four since single seriation, having attained a ceiling by wave four, was not assessed in wave five. The two sections are combined for correlational and related analyses which are independent of the means. To allow easier interpretation of these data and comparability from one variable to another, all means were standardised so that each measure across all years of assessment had a mean of 0 and a standard deviation of 1.0 before being plotted on graphs as shown in Figures 1 and 4.

PLACE FIGURE 1 ABOUT HERE

Considering increments in the means first, it is clear that for all four predictor variables, the years of the study have witnessed dramatic quantitative growth (Figure 1). Both Perceptual Regulation and Vocabulary follow a near linear function with little sign of any ceiling effect in the measures. Seriation (single and multiple 3 x 3) and Grammatic Prediction tend towards an asymptotic function with ceiling effects of the tests becoming evident at the fourth and fifth waves of measurement. However, when Seriation is expanded by adding the 5 x 5 matrix task, no ceiling effect is evident for the multiple seriation part of the test although a further year of testing would probably demonstrate the effect of a ceiling on that aspect of the measure also. As expected, all tasks show major increments in what Wohlwill (1973) calls "developmental function" during these first five years of

school. If only changes in the means were considered there is little or no difference in developmental pattern between the oral language and the operative variables.

However, when the stability values are considered in order to give an indication of individual differences from one year to another, there are striking contrasts. An inspection of the oral language stability values (underlined in Table 3) shows a uniform and strong pattern across the five years with very little difference between Grammatic Prediction and Vocabulary. In correlations of variables one year removed there are eight values which range from .60 to .75 with a mean of .66. These decline somewhat when variables measured two or more years apart are correlated but remain moderate to strong ranging from .40 to .68 with a mean of .53. This represents a high stability and indicates little change in the rank order from one year to the next and not very much more change across two, three, or even four years. Despite major increments in performance for the whole group, the order of students on these language capacities varies little during the years of the study, i.e., individual differences in development are not great.

The stability of the operative variables presents an entirely different picture. Perceptual Regulation, in all its first and second wave comparisons has low stability coefficients which range from .22 to .35 with a mean of .28 (Table 3). The strength of the stability estimate increases marginally from wave one to wave two probably due to an increase in the internal consistency of the measure recorded at the second testing. However, in comparisons with waves three and four there is a marked jump with the stability values rising to .65 and .66 when measures are compared one year apart and .45, two years apart. This difference cannot be explained by changes in reliability which does vary slightly but not in parallel with the stability values. If measures were taken only at waves one and two the interpretation would be difficult. Is

there a degree of developmental instability or is the measure a weak one? The high stability in relation to waves three and four rules out this second interpretation. The considerable switching in rank order across the first three measures reflects individual differences in development, suggesting that the first two years of school witness a period of marked developmental change most likely of a qualitative kind, that settles down in year three. The rank order established then is largely maintained for the next two years.

Seriation possesses moderate stability in all its comparisons across one year except in the last comparison, with coefficients of .46, .44, .46 and .28. This lower final value seems to be a measurement aberration as the test in wave five is shorter (with the first part, Single Seriation, omitted), captures less variance (especially in the second part, the matrix 3 x 3) and possesses the lowest alpha reliability (.59) for any of the variables over all waves of testing. The stability values across two or more years are low to moderate and range from .25 to .45 with a mean of .36. These values seem to reflect fairly considerable changes in rank order over the five years of testing. Changing the content of the test at year three (adding a 5 x 5 matrix task) to ensure appropriate measurement over the five year period seems to have largely sustained the variance of the test (despite some decline) as well as the level of individual differences across the five years. There seems to be no direct way to register the attainment of high stability on this kind of infralogical task since, when all subjects have mastered the task, the variance is small and correlations decline. Only by indirect assessment can the change from lower to higher stability be shown, for example, in a capacity which possesses some continuity but shows a disturbance in performance by lower stability for a period, as was found by McCall, Eichorn and Hogarty (1977) in the reanalysis of longitudinal IQ data and as has been found here for Perceptual Regulation.

In sum, the variables all reveal marked increases in developmental function (mean increments) over the period studied but varied levels of individual differences in this development as shown by the contrasting patterns of stability. This suggests that different kinds of developmental change may be taking place and, in accordance with the predictions based on theoretical grounds, two kinds of development may be detected. First, oral language shows quantitative increases and high stability implying minimal interindividual variation in development. Second, the two operative variables also increase quantitatively but overall possess far less stability suggesting marked individual differences as shown by a reshuffling of the rank order. After early instability, Perceptual Regulation settles down at the third wave and remains as stable as the language variables after that. Seriation largely reveals moderate levels of stability throughout because the variance has been maintained by deliberately modifying the instrument over the five years to allow for predicted growth. For both operative variables, this lower stability seems to imply that, unlike language, a qualitative change is occurring which produces these individual differences in development.

A further test of the nature of this change is now proposed by subjecting the predictor variable groupings to path analysis in order to study the relationship of the same grouping over the five testing occasions. Will this difference in the pattern of development be evident in the causal relationships of each of these two predictor variable groupings? The analysis, performed with the COSAN program (McDonald, 1980), provides a residual matrix to guide in the addition or subtraction of specific paths and a chi-square test to indicate overall goodness of fit.

The path analysis of language, shown in Figure 2, began with a simplex in which only direct links between successive years of measurement were postulated (Figure 2A). This structure did not reproduce the pattern of

correlations (χ^2 = 41.8, df = 7, p< .0001) so additional paths were inserted until an adequate model (Figure 2B) was derived (χ^2 = 1.7, df = 4, p = .78). This structure, which represents an auto-regressive process of order two (Anderson, 1971), displays a symmetrical pattern of development in which each language result, except the first, depends on two others, the measure immediately prior to it and the one before that. The variance explained at the four waves after the first, ranges from 52% to 64% with an average of 56%. The structure seems to represent a cumulative learning model in which each wave of longitudinal measurement exerts two years of direct influence on future development but its effect is then subsumed by later measures in successive waves. There seems to be a highly stable and regular pattern of change similar to the result indicated by the multitrait-multitime developmental analysis of the language variables.

PLACE FIGURE 2 ABOUT HERE

A quite different pattern emerged when operativity was analysed in a similar manner. The simplex (Figure 3A), in which only direct paths from one wave to the next were inserted, proved to be an inadequate fit (χ^2 = 41.8, df = 7, p < .0001) and more paths were added until the most parsimonious structure (Figure 3B) was attained (χ^2 = 2.92, df = 3, p = .40) This displays an asymmetrical developmental pattern in which kindergarten operativity continues to have a direct effect on each subsequent wave of operativity in addition to its indirect effects. Grade 2 operativity (wave three) has an additional direct effect on the Grade 4 capacity but this seems to be a result of the change in test content from wave three onward. The variance controlled at each point after the first is much lower than that controlled by language, having a range from 20% to 41% and an average of 29%.

PLACE FIGURE 3 ABOUT HERE

Kindergarten operativity seems to exert a unique effect across the whole range of measurement waves that declines as it becomes further removed but continues to be significant to the fifth wave. By contrast with the cumulative learning model of the language variables, there seems to be a basic process laid out in the first year (Kindergarten) which remains directly implicated in the development of operativity at each successive year. Certainly the path analysis presents further evidence that operativity and language follow quite different developmental patterns. The asynchronous pattern of operativity seems to give additional indirect evidence for a qualitative disturbance in development as noted in the previous analysis and as reported in other studies of this five to nine years age span (White, 1965, 1970).

Development and Construct Validity We are now in a position to reconsider the question of the construct validity of the operative variables discussed earlier, especially the apparently poor discriminant validity of Perceptual Regulation found by comparisons across the first three waves. The developmental patterns which have emerged in support of theoretical expectations change the interpretation of numerical values in respect of validity. If a variable is predicted to show instability during a certain period, low stability values are an index of sound validity if other explanations of those low values can be ruled out. Since validity seeks to determine whether we are measuring what we think we are measuring, as Cronbach and Meehl (1955) have asserted, the meaning of high or low stability will ultimately depend on theoretical assumptions. The study therefore requires

some reconsideration of the construct validity expectations discussed by Campbell and Fiske. These are based on psychometric assumptions that do not necessarily hold in some kinds of data (Wohlwill, in Kagan, 1980). It may be wondered whether convergent and discriminant validity apply at all to data which reveal developmental disturbance, presumably qualitative change, as is being claimed for some variables in this study. Yet, despite developmental discontinuity, continuity is also evident and the question must become one of whether a variable, even though it shows qualitative disturbance, maintains its identity sufficiently to be recognisable in terms of its original conception. This means that less stringent or different kinds of psychometric demands can be made on it to establish validity during a period of its developmental transformation. Thus, although the construct validity of the operativity variables appears to be considerably less strong than that of the oral language variables, it is nevertheless sufficient and a lower validity status does not seem appropriate. Oral language may simply suit the psychometric assumptions of the technique more readily.

On this basis there may be no need for concern about the validity of the operativity variable grouping when some of their early correlations are not significant and do not discriminate them well from other measures. As the data show, they do demonstrate a considerably greater convergence in correlations at the third and later waves. This is fully consonant with other findings that there is little inter relationship among operative capacities early in their emergence but a considerable relationship as they become established in the child's repertoire and the cognitive structures become integrated (Arlin, 1981; Flavell, 1971; Tomlinson-Keasey et al, 1979).

Development within the Criterion Variables. The results in regard to developmental change for the criterion variables are presented in three forms; mean values, stability patterns and path analysis of effects. The means

(Table 5) which were standardised and represented graphically in Figure 4,

PLACE FIGURE 4 ABOUT HERE

indicate marked increases in developmental function over the four years of testing, with Word Recognition following an almost perfect linear function and Paragraph Understanding showing a slight asymptotic tendency at the final wave.

The strong and consistent stability of Word Recognition (mean value .82) over this period shows that, despite dramatic change in developmental function, individual differences from one year to the next are slight. By contrast, Paragraph Understanding, although possessing moderate to strong stability with a mean of .54, is considerably less stable during these early years of school. It might be considered that this is caused by the presence of greater measurement error in the tests, however, an inspection of the reliability values (in parentheses, Table 4) does not support this explanation. There is little difference in the reliabilities of the two tests in waves two and three and a marginal advantage for Word Recognition in waves four and five. Perhaps by its nature a test of reading comprehension (Paragraph Understanding) is subject to greater developmental disturbance than is a test of word recognition in a similar way that operative variables display higher levels of individual differences in development over this period than do language variables. Furthermore, perhaps there is a stronger relationship between operativity and Paragraph Understanding than between operativity and Word Recognition and conversely a stronger relationship between the language variables and Word Recognition than exists between language and Paragraph Understanding. This possibility will be considered as relationships between predictor and criterion variables are analysed.

PLACE FIGURE 5 ABOUT HERE

The path analysis of the reading variable grouping, which is shown in Figure 5, was begun with the simplex (Figure 5A) which was, however, unable to reproduce the correlations, (χ^2 = 21.7, df = 4, p< .0002). An additional path was necessary between R2 and R4 (Figure 5B) to give an adequate representation (χ^2 = 1.87, df = 3, p = 0.6). This has the effect of reducing the coefficient of the direct link between R3 on R4 from .81 to .46. Recomposition of the effects at each year after the first indicates that the variance controlled by reading is greater than that controlled by either operativity or language and ranges from 65% to 72% with a mean of 69%. In sum, reading at Grades 2 and 3 depends on achievement at the previous grade in each case (Grades 1 and 2 respectively) but reading at Grade 4 is influenced by what is learnt at the two prior grades.

Relationship of Predictor to Criterion Variables

The results presented so far have demonstrated that the construct validity of the variables under study is satisfactory. In regard to the validity of their grouping, on multitrait-multitime assessment it is generally satisfactory but the operative variables show some fluctuation. This occurs, however, according to developmental expectations and seems to confirm further the validity of the grouping. Though all variables show a marked increase in mean values over the four years of the study, the lower stability of the operative variables suggests that they undergo a period of developmental surgence and a qualitative change which the language variables do not. This leads to the central issue of the study which is to examine the nature of the relationship

between operativity and reading and thus to test a cognitive developmental theory of learning to read. The study will now seek to follow this developmental pattern of operativity in its relationship to the child's emerging reading competence during the first five years of school in order to determine whether it has a causal effect upon that capacity. Its influence on reading will be compared with that of language which has been shown to follow a markedly different developmental progression.

Do the different developmental patterns which have been observed for operativity and oral language have distinct and different effects on the development of reading?

The relationship will be studied initially by examination of the simple correlations and of regression analysis. To explicate the pattern of effects for the purpose of causal attributions, path analysis will be used.

Prediction of Dependent Variables. The simple correlations of the operative variables, Perceptual Regulation and Seriation, and the language variables, Grammatic Prediction and Vocabulary, with the reading tasks, Word Recognition and Paragraph Understanding, are presented in Table 6.

The values are all positive, mostly significant, and low to moderate in strength. Seriation is generally more strongly correlated with reading than is Perceptual Regulation and Grammatic Prediction is more strongly correlated with reading than is Vocabulary. There is some change in the strength of the relationships over the years but no dramatic shifts. The language variables increase their relationship to reading with time although Vocabulary has the lowest correlations of all predictor variables and Grammatic Prediction possesses the highest. The operative variables seem more strongly related to

PLACE TABLE 6 ABOUT HERE

reading comprehension than to word reading and, to test this possibility, which has been suggested by the comparison of their stability patterns earlier, the operative variables were regressed onto Word Recognition and Paragraph Understanding in turn (Table 7). To test the converse, that language is more strongly related to word reading than to reading comprehension, a similar procedure was followed for the language variables (Table 7). This analysis gives further opportunity to compare the predictive patterns of operativity and language to reading and especially to examine the extended predictive relationships over the five annual testing occasions.

The multiple regression values of operativity with the two reading tests show that operativity is more closely related to reading comprehension than it is to word reading (Table 7). Of the 14 comparisons the correlation is stronger with Paragraph Understanding on 12 occasions. This cannot be explained as the result of less measurement error in the Paragraph Understanding Test since in all 12 occasions when operativity is more strongly related to Paragraph Understanding its reliability, being equal to or lower than that of Word Recognition, would bias the results in the opposite direction (Table 4). The two times when operativity is more strongly related to Word Recognition occur in comparisons with the first and second testing of reading when it is at its most elementary level and reading comprehension is in a relatively unformed state.

Thus one reason that reading comprehension is less stable than word recognition, as observed earlier, seems to be that it draws to a greater extent on the developmentally surgent operative capacities. The converse, that oral language is more closely related to word reading, does not gain support in the data (Table 7) since, of the 14 multiple correlation

comparisons, language is more strongly related to Word Recognition on six occasions and to Paragraph Understanding on eight. Its greater stability does not seem to stem from its stronger relationship to the more stable language variables than Paragraph Understanding.

Examination of the patterns of relationship across time for language to either reading measure and for operativity to word recognition reveals some fluctuation but no major changes (Table 7). By contrast there is generally an increase in the prediction of reading comprehension by operativity as the time interval between them increases. This is most distinct for wave one operativity with values rising from .33 at waves two and three, to .40 at wave four and .44 at wave five. This cannot be explained by increased reliability since the values do not follow this trend for either operativity (Table 3) or Paragraph Understanding (Table 4). This result, however, tends to parallel the pattern of increasing stability at least for the Perceptual Regulation component of operativity and for Paragraph Understanding. As has been discussed, these seem to reflect a qualitative change in development which may also underlie the increasing influence of operativity on reading comprehension over time. Further inspection of the data (Table 7) shows that, when concurrent multiple correlations with reading (of variables measured in the same year) are compared with distal correlations (of variable measured one or more years apart), a striking difference between the pattern of operativity and oral language associations is evident. In the prediction of reading comprehension or word recognition by oral language, concurrent correlations are generally stronger (in 15 out of 20 cases) as might be expected. In the prediction of reading by operativity, the opposite pattern exists; distal multiple correlations are generally stronger than concurrent ones (in 17 out of 20 cases). This pattern is most striking for wave one and the effect is much less in waves three, four and five but there can be no doubt about this unusual general trend in the data. Operativity measured a year or more previously in this age range predicts reading more strongly than does a

concurrent operativity measure. While there may be some influence of changing reliability on these results, the reliability pattern over the five years does not support this interpretation of the finding. Operativity, especially that measured at Kindergarten, seems to exert a unique influence on reading which will be further explored with path analysis.

In further considering the pattern of relationships, operativity at Kindergarten (wave 1) is, in all 8 comparisons, the stronger predictor of both reading measures. From Grade 1 onward (wave 2), of 20 comparisons, oral language is the stronger predictor in all except one. This seems to suggest a change in the causal relationships over time which will be further analysed,

PLACE TABLE 8 ABOUT HERE

first with a backward multiple regression procedure to test the relative strength of operative and other predictor variables, and second, with path analysis.

The multiple regression of the two operativity measures and the two oral language measures on a composite reading score (a standardised unweighted average of Word Recognition and Paragraph Understanding) and the effects of omitting each predictor from the equation are shown in Table 8. In all cases both operativity and language add significantly to the prediction of reading. None of the reductions in the coefficient caused by omitting a variable group fails to attain significance. A closer examination of the values indicates an important pattern in the data which confirms the impression gained from the correlational matrix (Table 6). At wave one, operativity is by far the more dominant variable with its omission causing the greater drop in the multiple correlation, e.g., from .45 to .30 for the first comparison while omission of language reduces the value to .40. However, from wave two onward, language

becomes the consistently stronger influence. This may have causal implications which will be further elucidated by path analysis.

Awareness of the units of print, which has been studied in detail as a criterion measure (Watson, 1984) is used in this study as a predictor variable of reading. Its relationship to reading when combined with the other predictors in waves two, three and four is shown in Table 9.

PLACE TABLE 9 ABOUT HERE

Its effects are clearly most powerful in its wave two measurement and less so but still significant in two out of three regressions in its wave three relationships. By wave four this measure exerts no significant influence. This pattern is not surprising since, although being able to discriminate a number, a letter, a word and a sentence is a task which is still being mastered in Grade 1, it is largely known in Grade 2 and finds very few who have not mastered it by Grade 3. The declining variance of the test (Table 5) bears this out. It seems that other metalinguistic tasks are required to discriminate amongst these older subjects on linguistic awareness.

Thus, analysis of the prediction of reading using simple correlations and multiple regression indicates that operativity and language are partly distinct from each other in their prediction of reading. In addition, operativity has a unique and dominant early relationship with reading (at Kindergarten) while oral language has the stronger relationship later (Grades 1, 2, 3 and 4). This is in keeping with the developmental pattern found earlier for each separately; language develops in a synchronous way and operativity in an asynchronous manner with a basic process laid out by Kindergarten influencing all later levels. This pattern of results suggests that operativity and oral language exert different causal effects on reading. Path analysis will now be used to elucidate this suggestion.

Pattern of Effects. The path analysis, which examines the causal relationships of the data across the five testing periods, has two major analytic phases; the first to establish the pattern of effects of operativity and language on reading and the second to assess the additional influence upon this pattern of an understanding of the units of print. This will allow the longitudinal testing of the hypothesis of the study that operativity has a causal influence on learning to read. The various path analytic models will be introduced in increasing order of complexity.

The first phase, linking operativity, language and reading, is addressed by the path diagram shown in Figure 6.

This structure is formed by linking the simplex diagrams from the three earlier analyses of language, operativity and reading (Figures 2A, 3A and 5A). The direction of effect is shown by the single headed arrows and the strength of effect by the path coefficients on the shaft of each arrow. A double headed arrow joins the two exogenous variables, Kindergarten operativity (OpK) and Kindergarten oral language (LK), for which there is no path coefficient since their causes come from outside the model and are not analysed in this study.

PLACE FIGURE 6 ABOUT HERE

An assumption of the analysis is that reading at any grade is influenced by operativity and language at the same grade and that Grade 1 reading (R1) is also influenced by Kindergarten operativity (OpK) and oral language (LK). This assumed direction of causal effect seems well justified during the first and second years of formal school instruction when the teaching of reading is only beginning and the children's reading ability is still largely unformed. To ensure that reading was not developed to any significant degree in the

sample prior to formal instruction, a word recognition test (St. Lucia, Andrews, 1969) was administered to subjects before they entered Grade 1. Although there were discernable differences depending on whether the Kindergarten had followed a subskills or an object-picture labelling program, only 5 of the 148 subjects recognised 10 or more words, over one third scored 0 or 1 and the mean was 3. This gives support for the assumption that reading does not exert any significant influence on at least the Kindergarten and Grade 1 operativity and oral language responses. It is reasonable that the more fundamental abilities, operativity and oral language, which are less reliant upon deliberate cultural initiation than reading, possess temporal and causal precedence over reading at least in Grades 1 and 2. As early reading skills are mastered and a degree of independence is attained, perhaps to some extent, in Grades 3 and 4, there is greater likelihood that it will begin to exert a causal influence on operativity and oral language -- a possibility which needs to be taken into account in the analysis of the later part of the longitudinal data.

At this stage of the analysis no links are inserted between operativity and language nor between reading at one grade and operativity or language at a subsequent grade. The previous path analysis of a part of the data (Figure 2) established that, while there were paths between Kindergarten operativity and both reading and language at Grade 1, no paths were needed between Kindergarten language and either Grade 1 reading or operativity. Thus the present structure was adjusted by the insertion of the paths required (e.g., OpK - R1, OpK - L1). This initial structure proved to be too simple to allow a good fit (χ^2 = 192.0, df = 70, $p < .0000$) so the additional paths determined earlier in the separate analyses of operativity, reading and language (Figures 2B, 3B, 5B) were inserted (Figure 6B).

This improved the fit (χ^2 = 96.3, df = 62, $p < .003$) but did not substantially change the basic relationships under study between operativity and reading and between language and reading. The relationships between the five measures of operativity at different grades are not affected since these can be regarded as exogenous variables in respect to reading and no links have been inserted between language and operativity. Similarly the path coefficients for the language aspect of the structure are unaltered from those found previously. Examination of the residuals (See Appendix A for residual table) suggest that further direct paths may need to be inserted between operativity and language (of the form Lj - Opj + 1 and Opj - Lj + 1) and between reading and language (Rj - Lj + 1) but that these will not significantly affect the path coefficients of central interest to the purposes of the analysis, i.e., between operativity and reading and language and reading.

The direct effects of operativity are strongest in the early waves, at Kindergarten and Grade 1 (.16, .23), but decreased quickly at Grade 2 and remain low at subsequent grades (.08, .08, .08). These three paths may well be eliminated without significantly weakening the model and further analysis will test this possibility.

By contrast, the effect of language does not diminish to the same extent. Although initial language (LK) has no significant direct effect on reading at Grade 1 (R1), the effect of language on reading at the same grade, despite some fluctuation, remains reasonably strong (.30, .15, .08, .16) over the four year period.

PLACE TABLE 10 ABOUT HERE

Introduction of the links between operativity and reading and between language and reading has the effect of reducing the coefficients between the four observations of reading achievement as a comparison of Figures 5A and 6B reveals.

At each stage after the first, the structure suggests a relationship of the form,

$Rj = B_1L + B_2Op + B_3 Rj - 1,$

with the effect of operativity diminishing after Grade 1, that of language being variable but not diminishing to the same extent, and the influence of reading at prior grades being the most important.

At the first stage (relationships with Grade 1 reading), however, there is a different pattern, with operativity playing a prominant part. In order to show clearly the relationships which exist at this part of the structure, the associations of Kindergarten and Grade 1 operativity and language with Grade 1 reading are decomposed (Table 10) using Duncan's fundamental theorem (Wolfle, 1980). This analysis reveals the considerable influence of Kindergarten operativity on Grade 1 reading, as a direct effect, as an indirect effect through Grade 1 operativity and as a spurious effect in regard to both Grade 1 operativity and language. Kindergarten operativity's direct effect on Grade 1 reading is .16 and its indirect effect is .19, giving a total effect of .35 or 90% of the total association. While the direct effect of Grade 1 operativity is .23 the spurious effect of Kindergarten operativity is .14 or 35% of the total association. Thus the effects of operativity (Kindergarten and Grade 1) in the association of Grade 1 operativity and Grade 1 reading is .37 (.23 + .14), or 93% of the total association. Kindergarten operativity also exerts a significant influence on the association of Grade 1 language and Grade 1 reading as a spurious effect (.05 or 12.5% of the total association). If Kindergarten operativity were left out of the analysis undue importance would

have been attached to spurious effects of Grade 1 operativity and Grade 1 language.

In accord with the hypothesis of the study this analysis shows that operativity at Kindergarten and Grade 1 exerts a major causal influence on reading which is distinct from that of oral language and, in addition, influences reading through oral language to some degree. Language at Kindergarten has no direct effect on Grade 1 reading but a substantial indirect effect (.19) while language at Grade 1 has a considerable direct effect (.30). Thus it may be seen that both operativity and language in Kindergarten and Grade 1 independently effect early reading, with operativity in Kindergarten appearing to play a particularly important causal role. This analysis has not yet taken account of the possible effects of the child's awareness of the units of print which seems to play an important part in early reading and to be causally dependent on operativity and language but particularly on operativity. Does it exert a causal influence in addition to the effects of operativity and language? The next phase of the analysis will address this question.

As in the previous analysis, a simple structure was first studied utilising the relationships found separately for operativity, reading and language and inserting awareness of the units of print into the structure (Figure 7). The assumptions of this model are that units of print influences reading when both are assessed at the same time and that operativity and language, assessed in the same wave, influence units of print rather than the converse in each case (Figure 7A).

PLACE FIGURE 7 ABOUT HERE

This structure is not adequate (χ^2 = 270, df = 106, p< .0001) as only the simplex structures have been used for operativity, reading and language (the only direct links used are between one observation and a subsequent one). However, it is sufficient to show that print awareness has a strong direct effect on reading at Grade 1 (.31) but that its subsequent influence is negligible (.04 at grade 2 and -.00 at grade 3). Operativity affects print awareness at Grade 1 (.30) but its influence fades at higher grades (.09, .04). Predictably after Grade 1, a test of the capacity to distinguish a number, a letter, a word and a sentence is not a critical capacity for reading. This level of metalinguistic knowledge does not effectively discriminate amongst subjects as the reduced variance indicates (Table 5) and it seems that other more difficult aspects of language awareness need to be incorporated into the test at Grade 2 and beyond.

In summary, the effect of introducing print awareness into the model is to decrease the path coefficient for the direct link from Kindergarten operativity to Grade 1 reading from .16 to .10, from Grade 1 operativity to Grade 1 reading from .23 to .14 and from Grade 1 language to Grade 1 reading from .30 to .22. This confirms the findings of the earlier study (Watson, 1984) by showing that part of the effect on Grade 1 reading of operativity at Kindergarten and Grade 1 and of language at Grade 1 is exerted through print awareness. Yet awareness of the units of print exerts a major direct influence (.31) beyond the effects of operativity or oral language and thus adds importantly to the explanatory power of the model.

Introduction of the additional paths (Figure 7B) for each of the reading, language and operative parts of the structure (from Figures 2B, 3B and 5B) has the effect of making the model more nearly adequate (χ^2 = 167.0, df = 97, p< .0001) and with a χ^2/df ratio of 1.7 this seems to provide a satisfactory solution (Schmitt, 1978). The effect of print awareness on reading is not modified as a comparison of the relevant parts of Figure 7A and 7B shows.

In an attempt to find a more parsimonious model the structure in Figure 7B was modified first by eliminating the direct paths between print awareness at Grades 2 and 3 and reading at the same grades (Figure 8). The path coefficients, which were determined again, showed little change and the test statistic increased only slightly (χ^2 = 168.6, df = 99) showing that these two paths can be eliminated (χ^2 = 1.6, df = 2, p< .4). The direct influence

PLACE FIGURE 8 ABOUT HERE

of operativity at Grades 2, 3 and 4 on reading at those three grades also looked as if it could be eliminated and this was next assessed. The result of setting these paths to zero made only a slight difference to the test statistic (χ^2 = 176.1, df = 102) indicating that their omission in the more parsimonious model (Figure 8) is appropriate (χ^2 = 7.5, df = 3, p< .05).

In this form of the model the relationships with reading may be summarised as follows.

At Grade 1,

R1 = .31 UOP1 + .14 Op1 + .10 OpK + .22 L1.

At Grade 2,

R2 = .76 R1 + .17 L1.

At Grade 3,

R3 = .81 R2 + .09 L3.

At Grade 4,

R4 = .45 R3 + .35 R2 + .17 L4.

This means that after Grade 1, reading is influenced primarily by prior reading and, to a lesser extent, by oral language measured at the same grade. Although there are substantial correlations between operativity and reading at Grades 2, 3 and 4 (.33, .34, .35) these do not exert an effect upon reading above the indirect influence of Kindergarten and Grade 1 operativity through Grade 1 reading. In regard to the effect of operativity on reading after

Grade 1, nothing new is happening. By contrast, the changes in language ability continue to exert an influence on reading across all grades assessed. Thus it may be seen that the different developmental patterns found for operativity and oral language have distinct and differential effects on the development of reading. Operativity develops in a distinctive way at Kindergarten (and to a lesser extent at Grade 1) and exerts a unique effect on reading and on language from this point. Language develops cumulatively across the years monitored and exerts a continuing influence on reading as it does.

Summary of Results

The main findings, drawn together from this rather lengthy presentation of results, are as follows.

1. The construct validity of all predictor variables, as assessed with multitrait-multitime analysis, is satisfactory. The language variables, Vocabulary and Grammatic Prediction, possess much stronger validity than the operative variables, Seriation and Perceptual Regulation, on this analysis.

2. The reading measures, Word Recognition and Paragraph Understanding, are closely related in multitrait-multitime analysis and their grouping as a composite reading measure is supported. Word Recognition is highly stable both from one year to the next and across several years and is clearly distinct from other reading measures. Paragraph Understanding is less stable and less distinct showing more individual differences in development over the period studied.

3. In regard to the validity of the grouping of variables, the language pair receive strong support from the analysis. The operative variable grouping receives support which varies from weak in waves one and two to strong in

waves three, four and five. This difference does not seem to be adequately explained by reliability differences, by test content variations, by practice effects or by testing irregularities but most likely requires an explanation in terms of developmental changes.

4. Developmental change, as indexed by increments in the means and by stability values, reveals important differences between the language and operative variables. Although all means increase dramatically, the stability of the language variables is very strong showing little change in rank order across time. The operative variables possess lower stability overall with Seriation showing moderate stability and Perceptual Regulation low over waves one, two and three (in the order of .30), then higher stability over waves three, four and five (in the order of .65). Since variation in reliability values cannot account for the data, this implies that there are far greater individual differences in the development of operativity than of language during this period. Two different forms of developmental change are evident: One which is more stable and continuous (language) and the other which is more surgent and discontinuous. In Perceptual Regulation the continuity evident in the high stability across waves three, four and five is absent in waves one, two and three due, it seems, to a period of developmental disturbance. Seriation, being a task mastered by virtually all during this period, shows discrete change with moderate stability consonant again with a considerable degree of developmental disequilibrium.

5. Path analysis of language and operative groupings further indicates that the two are quite distinct in their development. Language shows a symmetrical cumulative pattern over the five waves of assessment while operativity shows an asymmetrical pattern with the basic capacity laid out at the first measurement influencing all later responses.

6. This different developmental pattern has important implications for the assessment of construct validity in longitudinal data. Less stringent psychometric assumptions apply to variables which are undergoing a period of developmental surgence. Thus the apparent lower construct validity of operativity on multitrait-multitime analysis and the lack of early convergence of the two aspects of operativity tested, is consistent with theoretical predictions and, in itself, gives support for their construct validity.

7. The two criterion measures, Word Recognition and Paragraph Understanding, possess developmental similarities and differences. While the means of each increase dramatically, Word Recognition is highly stable and Paragraph Understanding is moderate to high in stability. Measurement error as recorded by the alpha reliability cannot account for this difference and Paragraph Understanding seems to be subject to the effects of greater developmental disturbance than Word Recognition.

8. In the multiple regression of operativity on Word Recognition and on Paragraph Understanding as separate reading criteria, operativity is consistently stronger in its relationship with Paragraph Understanding than with Word Recognition. The lower stability (greater level of individual differences) in the development of Paragraph Understanding than in Word Recognition (findings 2 and 6) is consistent with its closer relationship to the developmentally surgent operative capacities.

9. In the prediction of reading by operativity, distal associations are mostly greater than proximal ones while language follows the expected pattern by predicting more strongly to reading in its proximal measures.

10. The multiple regression analysis shows there is generally an increase over time in the prediction of reading comprehension by operativity. This is especially so in the case of Kindergarten operativity which predicts Grade 4 reading comprehension more strongly than does the concurrent operative measures and exceeds all temporally closer operative predictions of Grade 4 reading comprehension.

11. Both operativity and oral language, in a backward regression procedure contribute separately to the prediction of composite reading scores in all years of the project. In the first wave (Kindergarten), operativity is the dominant partner while at subsequent waves, oral language becomes dominant.

12. On the basis of the path analysis, Kindergarten and Grade 1 operativity plays an important causal role in the developmental of reading, directly, and indirectly through units of print and oral language. Language, which is overshadowed by operativity in its wave one effects, becomes the stronger causal influence at later waves.

13. Awareness of the units of print at Grade 1 is an important link in the causal chain between operativity and reading and to a lesser extent between language and reading. It also exerts an additional causal effect of its own above the influence of both operativity and oral language.

DISCUSSION

A Developmental Contrast: Operativity and Oral Language

Several lines of evidence in the results converge to give support for the view that operativity and oral language develop in distinctly different ways as proposed by the first hypothesis. Of greatest importance is the contrast between the stability patterns over the five waves of testing; language variables are highly stable throughout while the operative variables show considerable change in rank order of subjects. This suggests that operativity is subject to a degree of developmental disturbance in middle childhood which oral language is not. The different intravariable patterns of effect which each shows in path analysis is a second line of evidence for their developmental difference. Operativity's asymmetric path is quite distinct from the highly symmetric path of oral language. Another test of the difference in development between operativity and oral language is provided by their effects on reading over the five waves of testing. At waves one and two operativity has a marked effect after which its direct influence ceases - a finding consistent with the view that operativity gains a special impetus between five and seven years during which are laid out basic processes critical for learning to read. Language exerts a direct effect across all four waves at which both it and reading were assessed, suggesting considerable continuity of development. A further line of evidence is found in the apparent stronger construct validity of the language variables on multitrait - multitime analysis and in the stronger validity which a similar analysis gives to the grouping of the language variables by comparison with the grouping of the operative variables. These lower validity values for the operative variables, which are not adequately explained by differences in reliability or by other task and testing factors, seem best understood as the effects of the greater developmental surgence of operativity in middle childhood.

Thus, the different stability patterns of operativity and oral language, their different five year intravariable effects, their differing influences on reading and their different construct and grouping validity values, taken together, provide strong evidence for the conclusion in support of hypothesis one, that operativity and language, as measured in this study, develop in quite distinct ways during early school years. Language seems to show stable, continuous, cumulative change while operativity seems to show a surge of discontinuous development at ages five, six and seven.

This conclusion is congruent with the findings of Bayley (1970), derived from a much broader data base, that there are multiple mental abilities with some, such as language, which show more continuous and stable development and others, such as reasoning, which show less stability and appear to be more closely bound to a stage progression such as that proposed by Piaget. The conclusion is also consistent with the results of McCall, Appelbaum and Hogarty (1973) who, in the analysis of longitudinal IQ data, found a major developmental inflection at 6 years. The operativity variables seem to be strongly influenced by the same developmental disturbance. A further parallel to this finding is the conclusion of White (1965, 1970) drawn from the analysis of a large variety of studies from differing theoretical perspectives, that the years five to seven witness a broad spectrum of behavioural changes which can be summed up as a transition from a relatively associative response capacity to a more cognitive one. The many changes reviewed include a marked growth in inferential capacity, a more systematic perceptual exploration, the onset of resistence to classical conditioning, the increased knowledge of left-right orientation, an increased predictability of IQ, and the spontaneous use of a planning capacity. White proposes that, to the earlier developed associative response layer, a higher level of response capacity is added - a more cognitive function usually ascribed to "symbolic" or "abstract" thought. The operativity capacities seem to be a central expression of this cognitive layer.

Stage Progression

The findings provide important insights into the developmental nature of the predictor constructs. What support do the data give for a stage developmental view? While Brainerd's concept of a "measurement sequence" (Brainerd, 1978) with its psychometric assumptions of developmental continuity may perhaps be applied to the language variables, it does not seem to fit the operativity variables very well. For example, the change in stability values from .30 to .65 for Perceptual Regulation cannot be adequately explained in this way. By contrast, the developmental continuum proposed by Baltes and Nesselroade (1979) seems better able to accommodate the results, with operativity (possessing more stage-like qualities) located towards the strong end of the continuum and oral language towards the weak end. In order to consider the stage development issue more closely the data will be applied to the three questions raised in the earlier discussion of the nature of stages. First, is the change abrupt or gradual? Second, is it qualitative as well as quantitative? Third, are the subsystems of the hypothesized structure interrelated and, if so, how?

In regard to the first question all variables show regular mean increases which suggest gradual development. However, the lower stability values for operativity imply that it is characterized by considerable development upheaval for perhaps two or three years. The presence of this influence is demonstrated most clearly for Perceptual Regulation where the discontinuity is observed in the change from low stability over waves one two and three to high stability over waves three, four and five. Whether this effect only begins at about five to six years of age as assessed in wave one of this study cannot be determined from the present data. Parallel evidence (White, 1965), however, and the change from a response which is dominated by field effects at five years to one which is able to transcend them at seven or eight years, as in

the present data and also in the findings of chapter 2, suggests that this is likely. Thus, language seems to follow a gradual developmental trend throughout the early school years and probably beyond. Operativity also changes gradually at one level but is nevertheless subject, during two or three years, to some less gradual special influence which seems to be part of its essence rather than a peripheral effect. When viewed across a life span this may even be described as an abrupt change but in the limited time context and with the data of this project, that description does not seem warranted. Whether this developmental disturbance marks off a qualitative change needs consideration and is the subject of the second question.

Unlike language, which seems best described as undergoing quantitative change, operativity possesses characteristics which seem to change qualitatively. The available evidence from this and other studies (e.g. Elkind, 1964b) suggests that early seriation (e.g., the correct ordering of groups of three or four of the ten rods) is attained on the basis of a perceptual strategy while fully developed seriation seems to be attained as a conceptual reasoning process with the child placing the rods in correct order from a mental construction of the parts and not requiring visual checking for correct placement (Watson, 1979). In the case of perceptual regulation, the transition seems to be from figurative perception in which the child's response is dominated by field effects to an operational response in which field effects are transcended. In the first case the child at four or five perceives the most salient part of an ambiguous field while in the second case the child of seven or eight also sees less salient possibilities in the same visual field. It is hard to account for these kinds of changes in cumulative terms alone. The response difference involves essential characteristics which change in their quality. Thus, in considering the first and second questions on the nature of change, operativity, as represented in this study, seems to undergo a gradual but marked qualitative change over two or three years. This

begins to look rather like a stage but before this conclusion can be drawn, the third and most difficult question, concerning the interrelatedness of the component parts, must be considered. A number of unrelated qualitative changes would not be described as a stage since some integration of the different parts to produce a co-ordinated whole must be evident.

The data are very limited for assessing this third question because only one of Piaget's infralogical tasks, seriation, has been assessed and the only operative comparison can be with a capacity, perceptual regulation, that Piaget has described as "partly isomorphic" with concrete operations. Furthermore, a consideration of the two tests shows that they possess major task dissimilarity. Although both are individually administered, Perceptual Regulation requires an oral response to a visual presentation (line drawing) while Seriation requires the physical manipulation of objects on a board. The task requirements of Perceptual Regulation seem much closer to those of the language measures which call for an oral response to a visual presentation (line drawing). The finding of a significant relationship between Seriation and Perceptual Regulation, which is, at later waves, quite distinct from Vocabulary and Grammatic Prediction, provides notable evidence for some underlying commonality. But the relationship, which varies across time and gains strength at ages seven, eight and nine, is not a simple one. The two aspects of operativity assessed bear little relationship early in their development but are more strongly related as they become better established and presumably as the underlying structures become more closely integrated and stable. Whether this integration is one of several "nodal" points (Wohlwill, 1973) or a part of some "final assembly" (Flavell, 1971) is not clear but there does seem to be a degree of consolidation similar to that reported in other recent studies (Tomlinson - Keasey et al, 1979; Arlin, 1981). This developing interdependence of quite different but conceptually related skills, which undergo qualitative change, is evidence consistent with a stage

conception of development. The stage relationship seems more like the family resemblance model of Feldman and Toulmin (1975) than a Piagetian logico - mathematical structure but the data do not allow a satisfactory test of this suggestion. The results do provide further reason to retain the stage concept as a plausible hypothesis and this analysis makes the difference between the developmental continuity of oral language and the discontinuity of operativity the more striking. The question which must now be addressed is the nature and implications of the relationship between operativity and reading.

The Relationship of Operativity and Reading

In accord with their different patterns of development, operativity and language influence reading in different ways, both, in the path analysis, exerting a causal effect. Operativity, which has shown a consistent, moderate correlation with reading in a large range of studies (Waller, 1977) does not seem to be simply an incidental correlate nor only a facilitator of learning to read but a causal influence on reading. However, the attribution of a causal relationship requires close scrutiny. In what sense may operativity be considered as a cause of learning to read?

The use of a longitudinal design in the present investigation enables the time precedence of operativity over reading, an important condition for establishing causality, to be studied. Specific reading instruction was begun subsequent to the kindergarten assessment of the predictor variables and, although different kinds of informal experience of reading may have had some prior influence, this is slight and seems unable to account for the effects of Kindergarten operativity or language. At the Kindergarten level the more fundamental cognitive and language processes are assumed to have temporal and causal precedence over the more culturally specific skill of reading. Reading instruction in itself is not accepted as a sufficient explanation of learning

to read and the wide individual differences amongst children who have had similar experience of reading instruction (same teacher, materials, methods and length of time) bear this out.

Two other possible underlying determinants were examined in this study, oral language and metalinguistic processes. The path analysis (Figure 7) shows that the major influence of operativity on reading at Grade 1 is independent of the effects of oral language. This is further clarified by the decomposition of associations (Table 10) which shows that the most powerful effect on Grade 1 reading is exerted by Kindergarten operativity. The possibility that linguistic awareness in the form of knowledge of units of print underlies both operativity and reading is addressed in Figure 7 and by another study (Watson, 1984). Although print awareness at Grade 1 is importantly involved in reading at that grade, this analysis and an extensive treatment of the issue (Watson, 1984) shows that it is dependent on operativity at Kindergarten and Grade 1 and could not therefore be an alternative explanation to the influence of operativity on reading. A further possibility is that IQ is the common underlying determiner of both. However, the analysis shows that operativity acts independently of an IQ measure commonly used with young children, the Peabody Picture Vocabulary Test, which was employed in the study as one index of language. Although other IQ measures may have stronger effects, it seems most unlikely that they would eliminate operativity. Furthermore, even if IQ is an important causal influence on reading, its lack of any underlying theoretical basis means that IQ can contribute very little if anything to the explanatory purpose of the present study.

Such competing views as those involving oral language, print awareness, reading instruction and IQ therefore, in and of themselves, seem unlikely to provide satisfactory alternate causal explanations of learning to read. At the same time, however, it must be recognized that no such elimination

procedure for establishing causal effects can ever be complete. The proposed causal relationship between operativity and reading, while receiving strong support from the path analysis, must remain tentative, providing a working hypothesis in need of further scrutiny.

The findings also need some further qualification in accord with the limitations of the middle class sample and the restricted range of operativity and oral language abilities monitored. The use of middle class subjects limits the generality of the findings. The sample of Lunzer, Dolan and Wilkinson, (1976), which included a privileged, an under-privileged and an intermediate group, demonstrated social class differences in the relationship of operativity and reading despite a highly significant relationship for the whole sample. The intermediate group showed no significant effects while the other groups did. If the intermediate group is closest to the middle class sample of the present study, stronger correlations for upper and lower socio-economic groups might be expected. However, the different social class patterns in England and Australia make such comparisons difficult if not misleading and it would seem important to replicate the findings with other Australian socio-economic groups.

The practical demands of extensive individual testing over five occasions meant that a restricted range of operational and oral language capacities was monitored. With a larger number of operational tasks, Lunzer Dolan and Wilkinson (1976) and Arlin (1981) each obtained a stronger multiple correlation with reading, and, although neither study used both the measures employed here, it is most likely that a wider range of tasks would have produced a stronger prediction than in the present study. In a similar manner, a wider range of oral language tasks is also likely to produce a stronger prediction. The size of correlations in this study is therefore not a very satisfactory guide to the strength of the relationships or to the

importance of the results. The qualitative differences between the developmental patterns of operativity and oral language over the five years and the difference in their effects on reading are the most important findings of the study which seem likely to hold up even if the predictive strength of each was increased.

In sum, at least for a middle class sample, operativity seems to be a necessary though not a sufficient condition for the development of significant reading competence. Oral language capability may also be necessary and some form of social initiation, most commonly school instruction, is apparently required as a precipitating factor. But what level of operativity is needed for reading? This is the first of two current problems in the literature that will now be considered.

Some Current Issues

A closer examination of the special influence of operativity on early reading gives insight into two problems in the research and theory of reading. First, the issue of whether there is a critical developmental level necessary for learning to read and second, why there is an increase in the prediction of reading with the passage of time.

Critical Level. If operativity is necessary for learning to read and it develops gradually over a period of two or three years is there some critical level of operativity required to support beginning reading instruction? This raises the issue of bridging the gap between "groupments" and behaviour (Tomlinson - Keasey et al, 1979; Arlin, 1981) which has produced considerable controversy (e.g., Gallagher, 1979) and which allows no easy resolution (Waller, 1977). During the same time that operativity is developing, reading typically is also undergoing gradual change and since it is well advanced by

the time high levels of operativity are attained, a high level of operativity cannot be a necessity for the early learning tasks of reading. The concurrence of cognitive and language development with learning to read has led to the suggestion that "reading is not inextricably connected to a particular stage of development of cognitive and language skills" (Doehring and Aulls, 1979, p.30) but rather that any such skill deficiencies will be made up as reading progresses. This may be partly the case but there seems to be important limitations on this process.

The difficulty of learning to read varies condiderably. Some aspects such as responding orally to the visual configuration of some words seem capable of being learnt at the associative level but other aspects, such as those requiring manipulation of units, seem to need the decision or operative zone of functioning. Some minimal level of operativity may be required in reading soon after beginning at least as it is usually taught in schools at present.

Support for this explanation is provided by the unusual finding of this study that operativity has distal associations with reading as strong as or stronger than proximal associations with reading. In other words, contrary to the normal pattern of results in correlational studies, earlier measures of operativity are as good or better predictors of reading than concurrent ones. This is no chance single occurrence but is a consistent pattern with few exceptions throughout the data. Variations in reliability as gauged by internal consistency cannot account for this result and, although the variance for one part of operativity, Seriation, is greatest earlier, the other part, Perceptual Regulation, shows very little change in variance but follows a similar trend in its correlations with reading. Better distal prediction indicates that the determinants of variance are not constant and implies that the structure of ability varies across time. This finding is consistent with the view that there is qualitative change in the functioning of the intellect.

It seems that there is a critical level of operational development needed for marked reading progress and each year finds more children past that level so that the prediction of reading by operativity decreases successively. Language, by contrast, has no such critical level since its proximal associations with reading are greater than its distal associations.

This finding is also borne out by the path analysis which shows that there are causal influences of operativity on reading at Kindergarten and Grade 1 then no further effects despite continuing significant correlations at Grades 2, 3 and 4. By contrast, again, language shows a steady pattern of effect from Grade 1 to Grade 4. An earlier study (Watson, 1984) of print awareness, operativity and reading further supports this contention. For example, in the cross-tabulation (Table 6-2, p.107) subjects who were low in operativity at Grade 1, were poor in units of print (and in reading). The longitudinal data show that they attained at least a moderate level of operativity before, or as they improved in units of print, and before they improved in reading. Perhaps some progress can be made at the associative level but it seems that the reading task soon calls for the cognitive or operational layer of response that shows an important developmental surge at five, six and seven years and that marks off a qualitative change in function.

In sum, the data support the view that some minimal critical level of operational development is called upon for the child to make extended progress in early reading. This is not to suggest that elaborate concrete operational structures are prerequisites of reading nor that reading itself does not support further development of operativity. What is maintained is that some basic operational structures precede certain critical aspects of reading such as the processing of units required in phonic synthesis and the elaborate seriation of meaning required by sentence and paragraph comprehension.

Prediction over Time. A surprising finding of the study is that multiple correlations of operativity with reading comprehension generally increase as the time interval between measurements becomes longer. This is in contrast to most predictor studies where the correlation declines as intervals between measurement increase. It is most striking for Kindergarten operativity but is also the trend for operativity at Grade 1 and Grade 2 (Table 7). Correlations of language to reading comprehension have some tendency to rise but overall remain fairly steady. This result is most unlikely to be the product of the particular tests chosen since a similar increased prediction over time is reported in three other reading studies which used different predictor and criterion variables (Butler, Marsh, Sheppard and Sheppard, 1981; Clarke, Bruininks and Glaman, 1978; Feshbeck, Adelman and Fuller, 1977). Perhaps there is some important common process underlying reading development in all studies. Four possible explanations of the finding are considered.

The first possibility is that less measurement error occured in later testing of reading. However, the small changes that do occur in the reliability values of Paragraph Understanding (Table 4) follow the opposite direction to the increase in this predictor correlation and could not explain the increase. A second explanation is that the rise in the multiple correlation is due to an increase in the variance of reading comprehension. This is lent some credence by the increase in the standard deviations of Paragraph Understanding across waves one to three (Table 5). However, this pattern does not hold up consistently. From wave four to five, when, for example, the correlation of Kindergarten operativity with reading comprehension rises, the standard deviation of the comprehension test falls. The correlations also fail to follow the pattern of variance in waves three and four and a similar discrepancy from the variance pattern is shown in the predictions of reading comprehension by the oral language measures (Table 7). Increases in the variances which may have some influence, do not therefore

seem to be a sufficient explanation of the stronger correlations with the passage of time.

A third possibility is that reading performance becomes more predictable over this formative period as its component capacities settle down and become better established in the response repertoire. That the prediction of reading comprehension by neither operativity nor language falls over time supports this and suggests that there is some common effect on both sets of relationships, namely, an increased predictability of the predictor as is suggested by Butler et al (1981). However, this cannot be the full explanation. Operativity and oral language are quite different in the pattern of their relationship to reading comprehension. There seems to be a special contribution to the phenomenon by operativity. While its correlations with reading comprehension rise, those of oral language are generally stable.

The fourth possibility is that operativity exerts a delayed influence or a sleeper effect on reading comprehension. In this view reading slowly takes on more of the decision oriented character of that earlier cognitive change. The developmental surge of operativity found at Kindergarten and Grade 1 has a slow cumulative effect on reading which increases to Grade 4. Unlike the externally operating sleeper effect of Kagan and Moss (1962) this effect seems to come from within the structure of ability itself. The rise in the correlations over time thus seems best explained as an interaction of the developmental patterns of operativity and reading comprehension and, if this is so, it provides a strong argument in support of the causal influence of operativity. The pervasive influence of the developmental surge registered by Kindergarten operativity may be the reason for the increasing correlations over time in the other studies cited. Some of the best predictors in the Clarke et al study (1978), e.g., visual sequencing, visual closure, could be descriptions of the operativity measures, Seriation and Perceptual Regulation, used in the present study. Clarke et al (1978) explain that the most powerful

predictor, number facility, involves many items requiring complex understandings such as "more", "less", "largest", "next to the last", "most of", all of which seem to be drawing upon the decision-making concrete operational capacities. Furthermore, the highest correlations of these predictors, as in the present study, are with reading comprehension rather than with word recognition or word attack. Similarly, the factors isolated from the Kindergarten screening test by Butler et al (1981), which predict reading (mainly comprehension) more strongly over time, could well reflect a conver- gence of the effects of the surgent cognitive abilities in both predictor and criterion measures. The first factor, including figure copying and completing, seems to call for considerable perceptual regulation. The second, activities using language, seems to be based largely on language awareness which has been shown (Watson, 1984) to be strongly dependent on operativity. If it is objected that operativity is not likely to affect all three factors isolated in the analysis, it must be remembered that even with the very narrow sampling of operative variables in this study, Kindergarten operativity exerts a highly pervasive effect (Figure 7), influencing the Grade 1 measures of language awareness, operativity, oral language and reading.

CONCLUSION

The results suggest that the operativity measures of this study are tapping a powerful developmental influence which is partly distinct from oral language in its effects on reading. This influence appears to be part of a broad spectrum of change in mental processing that occurs during middle childhood. This emerging conceptual capacity seems to be a necessary although not a sufficient condition for the development of reading ability. As well as exerting an immediate influence, operativity also produces a pervasive delayed effect in its causal relationship with reading comprehension which, during the early school years, slowly takes on the decision-making, conceptual reasoning character of that operational influence.

Table 1. Longitudinal Design of the Study

	Wave 1 Summer 1977	Wave 2 Summer 1978	Wave 3 Summer 1979	Wave 4 Summer 1980	Wave 5 Summer 1981
School Year	Kind.	Gr 1	Gr 2	Gr 3	Gr 4
N - core sample	148	142	136	135	128
N - control sample	-	-	44	-	-
Reading Measures					
Word Recognition		*	*	*	*
Paragraph Understanding		*	*	*	*
GAP (Cloze)				*	*
Progressive Achievement Comprehension					*
Piagetian Measures					
Seriation					
Single Seriation	*	*	*	*	
Multiple Ser. 3 x 3	*	*	*	*	*
Multiple Ser. 5 x 5			*	*	*
Perceptual Regulation	*	*	*	*	*
Language Measures					
Picture Vocabulary	*	*	*	*	*
Grammatic Closure	*	*	*	*	*
Units of Print Test		*	*	*	

Table 2. Effects of Repeated Testing Assessed at Wave Three for All Variables

	MEAN		Sig. of F for the difference
	Core Sample N = 136	Control Sample N = 44	
Word Recognition	34.07	34.64	-
Paragraph Understanding	12.84	11.16	-
Seriation	37.19	36.59	-
Perceptual Regulation	28.15	27.64	-
Vocabulary	71.32	74.96	x
Grammatic Prediction	28.15	27.64	-
Units of Print	19.69	19.93	-

x = $p < .05$

Table 3. Multitrait - Multitime Matrix of the Predictor Variables over Five Years (Waves)

	WAVE 1				WAVE 2				WAVE 3				WAVE 4				WAVE 5			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Wave 1																				
1. PR	(64)																			
2. Ser	10	(78)																		
3. GP	20	13	(97)																	
4. Voc	23	04	43	(90)																
Wave 2																				
1. PR	29	27	23	07	(79)															
2. Ser	17	46	13	10	27	(72)														
3. GP	28	34	69	41	31	33	(99)													
4. Voc	35	14	47	60	07	12	54	(99)												
Wave 3																				
1. PR	22	20	15	12	30	16	25	11	(86)											
2. Ser	14	47	15	04	16	44	26	13	23	(97)										
3. GP	19	28	60	31	20	27	68	44	17	26	(99)									
4. Voc	32	17	35	45	19	13	33	64	06	10	30	(94)								
Wave 4																				
1. PR	27	20	27	26	35	20	29	21	65	34	24	18	(77)							
2. Ser	13	39	04	03	01	28	08	-01	17	46	14	08	25	(86)						
3. GP	22	29	45	24	20	27	68	34	15	30	64	20	21	07	(89)					
4. Voc	26	25	42	48	21	18	44	63	15	16	42	61	25	09	30	(92)				
Wave 5																				
1. PR	23	22	24	26	33	19	28	17	47	30	23	15	66	18	15	27	(68)			
2. Ser	20	25	15	14	12	32	17	05	27	45	15	05	26	28	11	11	26	(59)		
3. GP	16	22	40	19	15	26	48	19	13	19	56	14	10	03	63	24	13	20	(64)	
4. Voc	29	21	43	45	25	20	46	60	13	15	45	56	30	08	41	75	35	02	29	(80)

PR = Perceptual Regulation
Ser = Seriation
GP = Grammatic Prediction
Voc = Vocabulary
Wave 1 = Kindergarten
Waves 2,3,4,5 = Grades 2,3,4 and 5 respectively.

Decimal points have been omitted for convenience of presentation.
Values in parentheses are reliabilities (Cronbach's alpha).
Values underlined constitute the stability value diagonals of each block.

Table 4. Multitrait-Multitime Matrix of the Reading Measures over Four Years.

		WAVE 2		WAVE 3		WAVE 4			WAVE 5			
		WR	PU	WR	PU	WR	PU	C1	WR	PU	C1	PA
WAVE 2	WR	(94)										
	PU	50	(96)									
WAVE 3	WR	85	52	(95)								
	PU	57	57	62	(95)							
WAVE 4	WR	77	47	84	60	(98)						
	PU	53	40	58	64	55	(90)					
	C1	55	44	62	59	62	66	(88)				
WAVE 5	WR	76	38	83	58	86	51	66	(92)			
	PU	58	42	64	57	58	63	66	67	(82)		
	C1	46	41	51	45	49	50	71	52	63	(84)	
	PR	44	40	57	52	52	60	67	59	66	68	(83)

WR - Word Recognition, St. Lucia Graded Word Reading Test
PU - Paragraph Understanding, Reading Comprehension
C1 - Cloze Reading
PA - Progressive Achievement Test, Reading Comprehension

Table 5 Means and Standard Deviations for All Measures

	WAVE 1		WAVE 2		WAVE 3		WAVE 4		WAVE 5	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
P R	8.19	3.04	10.40	3.16	11.67	3.05	13.49	3.48	14.92	3.37
Sr I	26.50	9.58	34.20	5.29	37.19	3.32	38.87	2.68	39.60	2.38
Sr II	-	-	-	-	31.05	4.42	32.73	3.70	35.24	3.02
GP	20.16	4.49	24.27	4.16	28.15	3.19	29.63	2.23	30.73	1.73
Voc	59.45	8.06	64.37	7.78	71.32	10.29	76.86	9.92	82.98	6.86
WR	-	-	21.25	8.33	34.07	9.57	44.44	12.34	54.21	12.15
PU	-	-	3.15	3.45	12.84	6.65	20.83	7.73	25.02	5.29
Clo	-	-	-	-	-	-	23.60	6.44	28.33	5.61
PAT	-	-	-	-	-	-	-	-	19.25	6.86
UOP	-	-	17.80	3.60	19.69	2.18	20.87	1.67	-	-

PR = Perceptual Regulation
Sr I = Seriation: Single & Multiple 3 x 3
Sr II = Seriation: Multiple 3 x 3 and 5 x 5
GP = Grammatic Prediction
Voc = Vocabulary
WR = Word Recognition
PU = Paragraph Understanding Test
Clo = Cloze: Reading
PAT = Progressive Achievement Test: Reading
UOP = Units of Print

Table 6. Correlation Matrix for the Prediction of Reading by Operativity and Language Variables

		WAVE 1		WAVE 2		WAVE 3		WAVE 4	
		WR	PU	WR	PU	WR	PU	WR	PU
Wave 1.	1.PR	.21	.20	.24	.19	.20	.18	.21	.31
	2.Ser	.31	.28	.33	.28	.34	.38	.27	.34
	3.GP	.28	.24	.27	.31	.27	.25	.27	.29
	4.Voc	.07	.20	.10	.09	.12	.15	.14	.14
Wave 2.	1.PR	.18	.26	.23	.30	.27	.30	.27	.30
	2.Ser	.28	.31	.22	.35	.17	.38	.22	.28
	3.GP	.46	.44	.47	.45	.44	.50	.43	.39
	4.Voc	.07	.25	.14	.18	.09	.24	.10	.18
Wave 3.	1.PR			.18	.21	.11	.25	.20	.30
	2.Ser			.26	.29	.23	.33	.26	.34
	3.GP			.39	.41	.38	.41	.45	.45
	4.Voc			.15	.21	.13	.21	.13	.22
Wave 4.	1.PR					.21	.35	.24	.29
	2.Ser					.22	.22	.20	.28
	3.GP					.47	.35	.48	.43
	4.Voc					.23	.30	.19	.24
Wave 5.	1.PR							.18	.28
	2.Ser							.24	.34
	3.GP							.45	.47
	4.Voc							.22	.25

WR = Word Recognition
PU = Paragraph Understanding Test
PR = Perceptual Regulation
Ser = Seriation
GP = Grammatic Prediction
Voc = Vocabulary

Table 7. Multiple Regression on the Operative and Oral Language Variables of Word Recognition and Paragraph Understanding

		WAVE 2		WAVE 3		WAVE 4		WAVE 5	
		WR	PU	WR	PU	WR	PU	WR	PU
Wave 1	Oper	.36	.33	.39	.33	.38	.40	.33	.44
	Lang	.28	.26	.27	.32	.27	.25	.27	.29
Wave 2	Oper	.31	.38	.29	.44	.30	.45	.32	.39
	Lang	.50	.44	.49	.46	.47	.50	.45	.39
Wave 3	Oper			.28	.32	.23	.37	.29	.41
	Lang			.39	.42	.38	.42	.45	.46
Wave 4	Oper					.27	.36	.28	.36
	Lang					.48	.41	.48	.44
Wave 5	Oper							.27	.39
	Lang							.46	.49

WR = Word Recognition
PU = Paragraph Understanding
Oper = Operativity
Lang = Oral Language

Table 8. Multiple Regression of Reading on Operativity and Oral Language with the Effects of Omitting Operativity and Oral Language in turn from the Equation.

		Wave 2 Reading Mult R	Sig	Wave 3 Reading Mult R	Sig	Wave 4 Reading Mult R	Sig	Wave 5 Reading Mult R	Sig
Wave 1	FM	.45	-	.47	-	.49	-	.47	-
	Omit Oper	.30	xx	.33	xx	.30	xx	.30	xx
	Omit Lang	.40	x	.39	xx	.44	x	.41	x
Wave 2	FM	.56	-	.56	-	.58	-	.50	-
	Omit Oper	.53	x	.53	x	.54	x	.45	x
	Omit Lang	.40	xx	.40	xx	.42	xx	.38	xx
Wave 3	FM			.50	-	.50	-	.55	-
	Omit Oper			.45	x	.45	x	.49	xx
	Omit Lang			.34	xx	.34	xx	.38	xx
Wave 4	FM					.55	-	.57	-
	Omit Oper					.50	x	.51	xx
	Omit Lang					.35	xx	.34	xx
Wave 5	FM							.57	-
	Omit Oper							.52	xx
	Omit Lang							.36	xx

Sig = significance of F for the omission of the variable grouping
xx = p .01, x = p .05
FM = Full Model
Oper = Operativity
Lang = Language

Table 9. Multiple Regression of Reading on Units of Print, Operativity and on Language with the Effects of Omitting Units of Print from the Equation

		Wave 2 Reading		Wave 3 Reading		Wave 4 Reading		Wave 5 Reading	
		R	Sig	R	Sig	R	Sig	R	Sig
Wave 2	FM	.62	-	.60	-	.62	-	.53	-
	Omit UOP	.56	xx	.56	xx	.58	xx	.50	
Wave 3	FM			.54	-	.54	-	.58	-
	Omit UOP			.50		.50	x	.55	x
Wave 4	FM					.56	-	.57	-
	Omit UOP					.55		.57	

Sig = Significance of F ratio for omitting UOP from the equation. xx = p .01 x = p .05

FM = Full Model (Operativity, Oral Language, Units of Print)

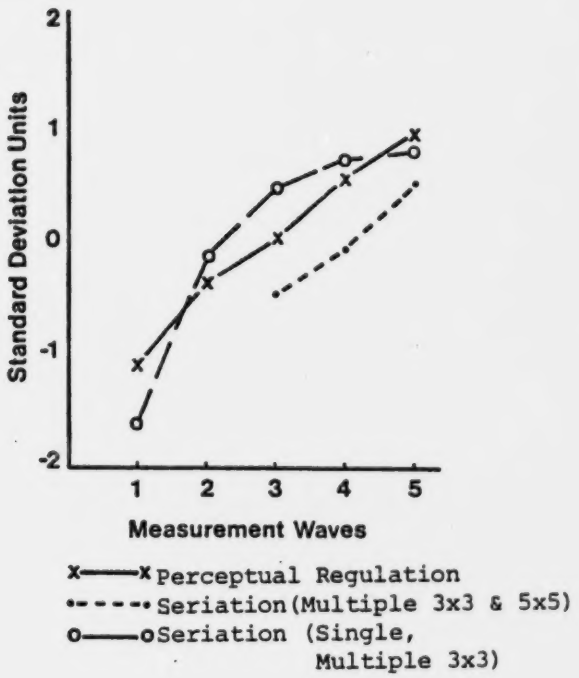
UOP = Units of Print

Table 10. Decomposition of Associations of Kindergarten and Grade 1 Operativity and Oral Language with Grade 1 Reading

VARIABLES	TYPE OF EFFECT	DECOMPOSITION		
Kindergarten Operativity & Grade 1 Reading	Total Association	.39		
	Total Effect		.35	
	Direct Effect			.16
	Indirect Effect			.19
	Through Op1			.07
	Through L1			.12
	Joint Association		.04	
Kindergarten Oral Language & Grade 1 Reading	Total Association	.27		
	Total Effect		.19	
	Direct Effect			--
	Indirect Effect			.19
	Through L1			.19
	Joint Association		.08	
Grade 1 Operativity and Grade 1 Reading	Total Association	.40		
	Total Effect		.23	
	Direct Effect			.23
	Indirect Effect			--
	Spurious Effects		.14	
	By OpK			.09
	By OpK through L1			.05
	Joint Association		.03	
Grade 1 Language and Grade 1 Reading	Total Association	.40		
	Total Effect		.30	
	Direct Effect			.30
	Indirect Effect			--
	Spurious Effects		.05	
	By OpK			.04
	By OpK through OpK			.01
	Joint Association		.05	

Figure 1. Standardized Mean Scores for Predictor Variables over Five Waves of Measurement (Kindergarten to Grade 4).

A. Operativity variables



B. Oral Language Variables

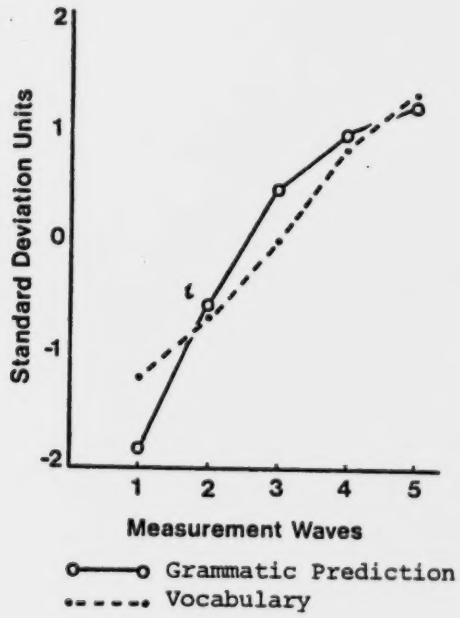


Figure 2. Path Analysis of the Relationships between Oral Language (Grammatic Closure and Vocabulary) over Five Waves of Measurement (Kindergarten to Grade 4).

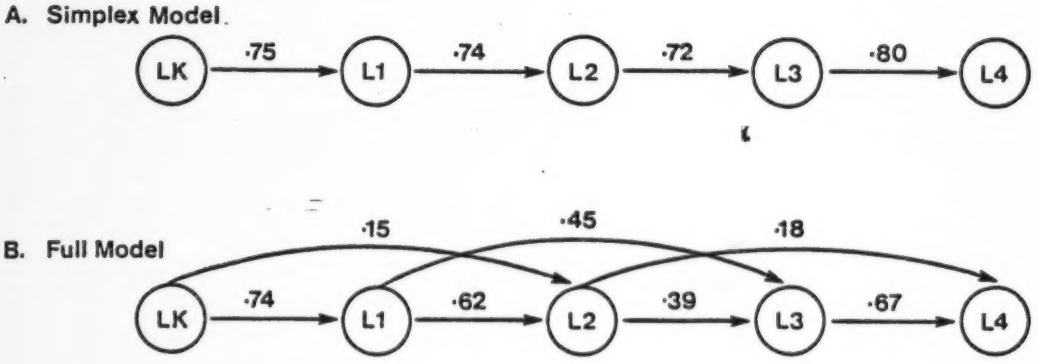


Figure 3. Path Analysis of the Relationships between Operativity (Perceptual Regulations and Seriation) over Five Waves of Measurement (Kindergarten to Grade 4).

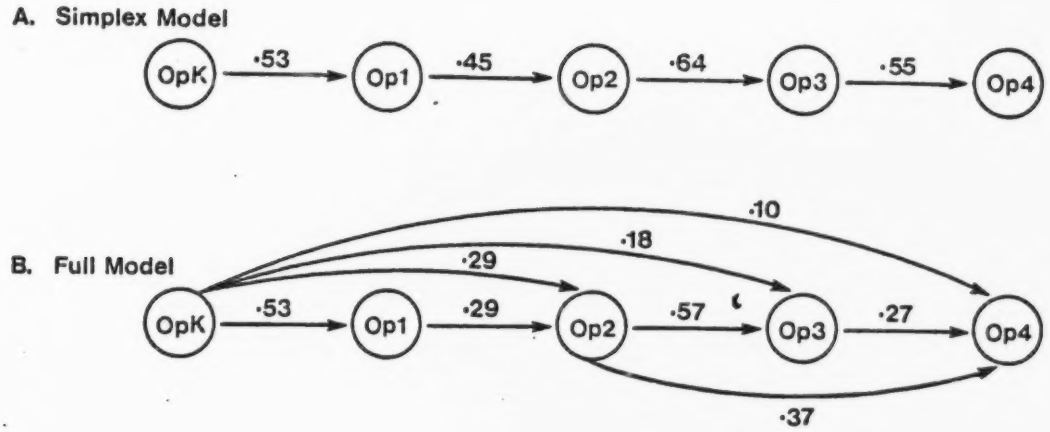


Figure 4. Standardized Mean Scores for Reading over Four Waves of Measurement (Grade 1 to Grade 4).

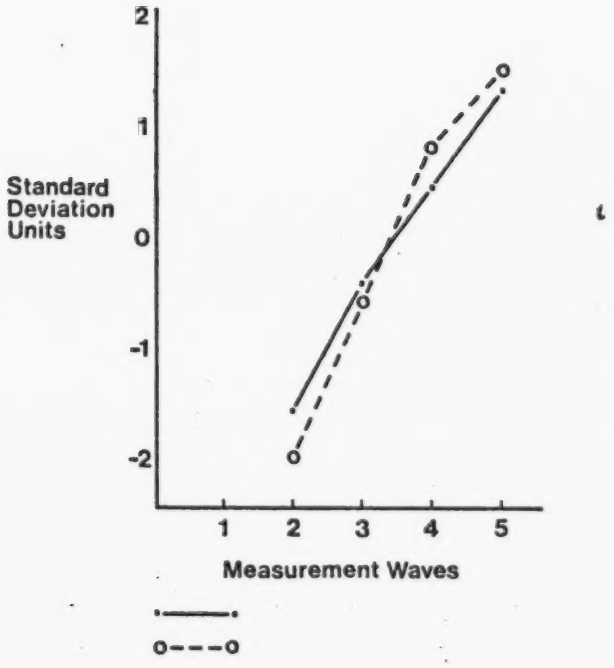


Figure 5. Path Analysis of the Relationships between Measures of Reading taken over Four Waves (Grade 1 to Grade 4).

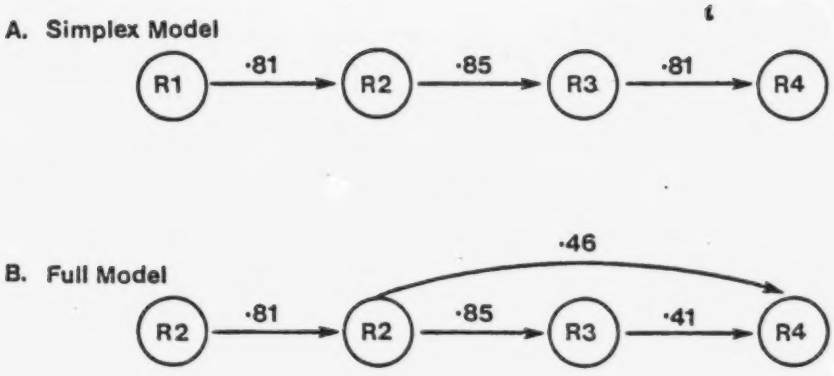


Figure 6. Path Analysis of the Relationships between measures of Operativity, Oral Language and Reading taken over five waves (Kindergarten to Grade 4).

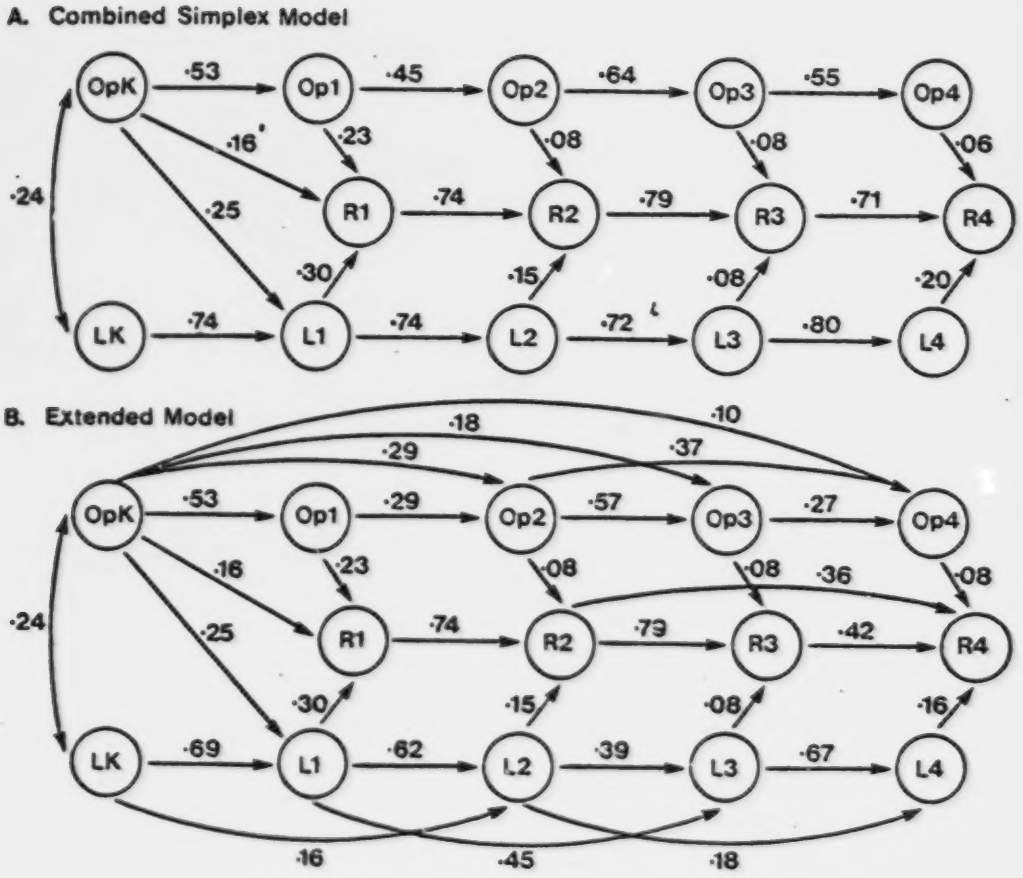
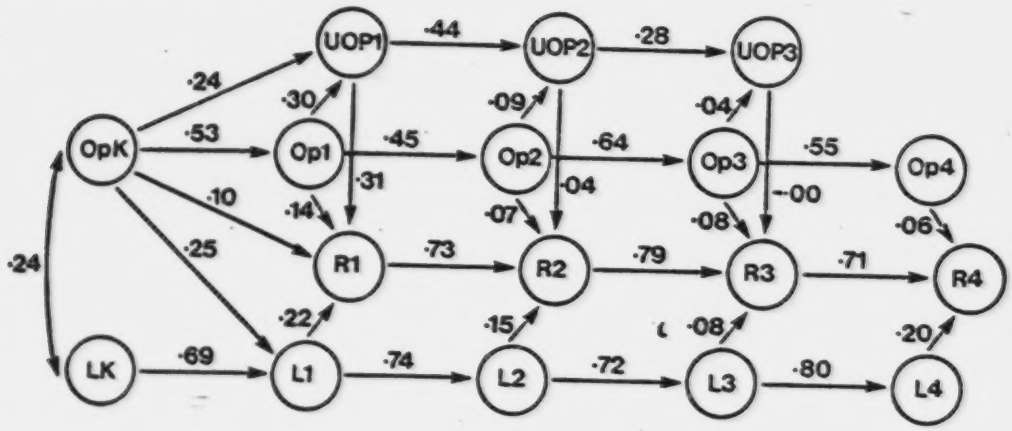


Figure 7. Path Analysis of the Relationships between Units of Print, Operativity, Oral Language and Reading over five waves of measurement (Kindergarten to Grade 4).

A. Combined Simplex Model



B. Extended Model

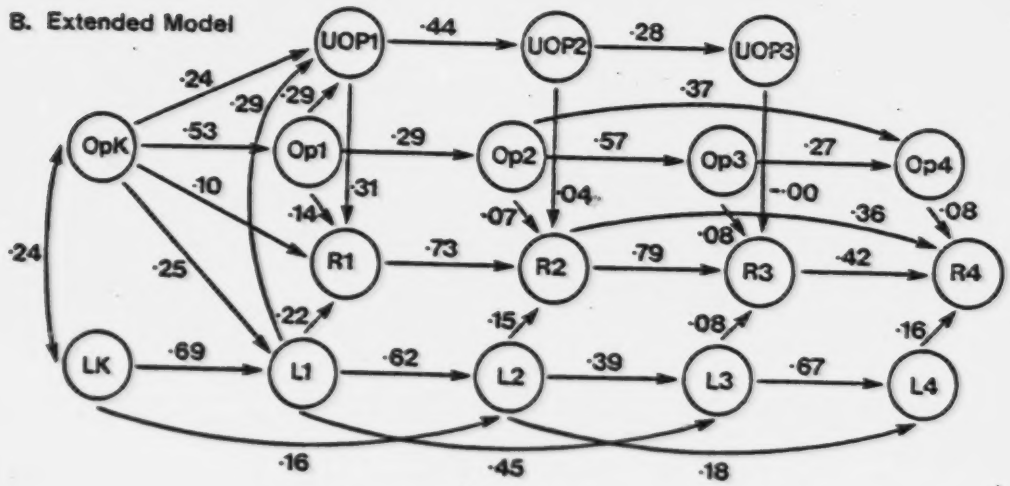
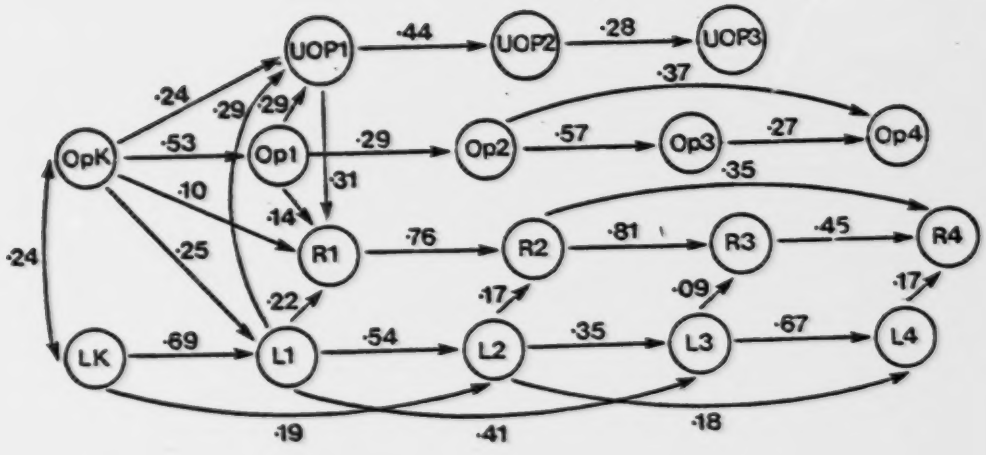


Figure 8. Path Analysis of the Relationships between Units of Print, Operativity, Oral Language and Reading over five waves of measurement (Kindergarten to Grade 4). Final Model.



APPENDIX A

Residual Matrix for the Path Analysis

Key to Variables Figure 6B

1	Op 1	Operativity Grade 1
2	Op 2	Operativity Grade 2
3	Op 3	Operativity Grade 3
4	Op 4	Operativity Grade 4
5	L 1	Oral Language Grade 1
6	L 2	Oral Language Grade 2
7	L 3	Oral Language Grade 3
8	L 4	Oral Language Grade 4
9	R 1	Reading Grade 1
10	R 2	Reading Grade 2
11	R 3	Reading Grade 3
12	R 4	Reading Grade 4
13	UOp1	Units of Print Grade 1
14	UOp2	Units of Print Grade 2
15	UOp3	Units of Print Grade 3
16	Op K	Operativity Kindergarten
17	L K	Oral Language Kingergarten

RESIDUAL MATRIX

		1	2	3	4	5	6	7	8	9	10	11	12	13
1	OP1	.0000												
2	OP2	.0000	.0000											
3	OP3	.0128	.0300	.0000										
4	OP4	.0929	.0000	.0000	.0000									
5	L1	.0899	.0830	.0230	.0823	.0000								
6	L2	.1707	.1022	.1255	.1159	.0000	.0000							
7	L3	.1953	.1657	.1129	.1389	-.0000	-.0000	-.0000						
8	L4	.2229	.1132	.0972	.1740	-.0050	.0000	-.0000	-.0000					
9	R1	.0238	.0699	.1520	.0548	.0209	.0126	.0981	.0785	.0108				
10	R2	.0789	.0673	.1035	.0735	-.0329	.0170	.1552	.1109	.0151	.0186			
11	R3	.1291	.0699	.0901	.1305	.0199	.0396	.1306	.0711	.0536	.0347	.0493		
12	R4	.1093	.1109	.1098	.1084	-.0836	.0338	.0867	.0829	.0674	.0494	.0522	.0594	
13	OPK	.0000	.0000	.0000	.0000	.0000	.1099	.1261	.1160	.0000	.0392	.0744	.0653	.0000
14	LK	.0749	.0490	.1287	.2057	.0000	.0000	.0071	.0426	.0170	-.0549	-.0128	-.0386	.0000

OPER LANG READ

RESIDUAL MATRIX

		14
14		.0000

DISTRIBUTION OF MAGNITUDES OF RESIDUALS

RANGE				COUNT			
GREATER THAN OR EQUAL TO	.200			2	=	1.9 %	} 50% > .2
GREATER THAN OR EQUAL TO	.100	BUT LESS THAN	.200	25	=	23.8 %	
GREATER THAN OR EQUAL TO	.050	BUT LESS THAN	.100	27	=	25.7 %	
GREATER THAN OR EQUAL TO	.010	BUT LESS THAN	.050	21	=	20.0 %	
GREATER THAN OR EQUAL TO	.005	BUT LESS THAN	.010	1	=	1.0 %	
GREATER THAN OR EQUAL TO	.001	BUT LESS THAN	.005	1	=	1.0 %	
		LESS THAN	.001	28	=	26.7 %	

MEAN = .05926407 LARGEST = .22241980 SMALLEST = .1735-017

KEY. OP1 – OP4 = Operativity, Grades 1 – 4. L1 – L4 Language, Grades 1 – 4.
R1 – R4 = Reading, Grades 1 – 4. OPK = Operativity Kindergarten. LK Language Kindergarten.

BEST COPY AVAILABLE

REFERENCES

Almy, M., Chittenden, E. & Miller, P. (1966). Young Children's Thinking: studies of some aspects of Piaget's theory. New York: Teachers College Press.

Almy, M., Dimitrovsky, L., Hardeman, M., Gordis, F., Chittenden, E., Elliott, D.L. (1970). Logical Thinking In Second Grade. New York: Teachers College Press.

Andrews, R. J. (1969). St. Lucia Graded Word Reading Test. Brisbane: Teaching & Testing Resources.

Arlin, P.K. (1981). Piagetian tasks as predictions of reading and math readiness in grades K - 1. Journal of Educational Psychology, 73, 712-721.

Ausubel, D.P., (1957). Theory and Problems of Child Development. New York: Grane and Stratton.

Baltes, P.B., Reese, H.W., & Lipsitt, L.P. (1980). Life-Span Developmental Psychology. Annual Review of Psychology, 31.

Baltes, P. B. and Nesselroade, J. R. (1979). History and rationale of longitudinal research. In J. R. Nesselroade and P. B. Baltes (Eds.), Longitudinal Research in the Study of Behaviour and Development. New York: Academic Press.

Bandura, A. (1971). Social Learning Theory. New York: General Learning Press.

Bayley, N. (1970). Development of mental abilities. In P. H. Mussen (Ed.), Carmichael's Manual of Child Psychology, Vol. 1. New York: Wiley.

Brainerd, C.J. (1978). The stage question in cognitive-developmental theory. The Behavioral and Brain Sciences. 2, 173-213.

Brainerd, C.J. and Allen, C.W. (1971). Experimental induction of the conservation of "first-order" quantitative invariants. Psychological Bulletin, 75, 128-144.

Butler, S.R., Marsh, H.W., Sheppard, M.J. & Sheppard, J.L. (1982). Early prediction of reading achievement with the Sheppard School Entry Screen Test: A four-year longitudinal study. Journal of Educational Psychology, 74 280-290.

Campbell, D. T. and Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait - multimethod matrix. Psychological Bulletin, 1959, 56, 81-105. Reprinted and abridged in D. A. Payne and R. F. McMorris (Eds.), Educational and Psychological Measurement. New Jersey: General.

Clark, C.R., Bruininks, R.H., & Glaman, G.U. (1978). Kindergarten predictors of three aspects of reading achievement. Perceptual and Motor Skills, 46, 411-419.

Clay, M.M. (1972). Reading: The Patterning of Complex Behaviour. Heinmann, second edition, 1979.

Danner, R. & Day, M. (1977). Eliciting formal operations. Child Development, 48, 1600-1606.

Denner, B. (1970). Representational and syntactic competence of problem readers. Child Development, 41, 881-882.

Dept. of Education, N.S.W. (1979). Reading K 12 Curriculum Policy Statement.

Dimitrovsky, L. & Almy, M. (1975). Early Conservation as a predictor of later reading. Journal of Psychology, 90, 11-18.

Doehring, D.G. & Aulls, M. W. (1979). The interactive nature of reading acquisition. Journal of Reading Behavior, 11, 27-39.

Donaldson, M. (1978). Children's Minds. Glasgow: Fontana.

Duncan, O.D. (1975). Introduction to Structural Equation Models. New York: Academic.

Dunn, L. M. (1959). Peabody Picture Vocabulary Test. Minnesota: American Guidance Service.

Ehri, L. C. (1979). Linguistic insight: Threshold of reading acquisition. In T. G. Waller and G. E. Mackinnon (Eds.), Reading Research: Advances in Theory and Practice, Volume I. New York: Academic Press. 63-111.

Elkind, D. (1975). Perceptual development in children. American Scientist, 63, 533-541.

Elkind, D. and Deblinger, J.A. (1969). Perceptual training achievement in disadvantaged children. Child Development, 40, 11-19.

Elkind, D., Horn, J & Schneider, G. (1965). Modified word recognition, reading achievement and perceptual decentration. Journal of Genetic Psychology, 107, 235-251.

Elkind, D., Larson, M. and Van Doorninck. (1965). Perceptual decentration learning and performance in slow and average readers. Journal of Educational Psychology, 56, 50-56.

Elkind, D. and Scott, L. (1962). Studies in perceptual development I: The decentration of perception. Child Development, 33, 619-630.

Elley, W.B. and Reid, N. (1969). The Progressive Achievement Tests of Reading Comprehension and Vocabulary. N.Z. Council for Education Research. Australian Edition, M.L. Clark, Hawthorn: A.C.E.R., 1973.

Feldman, C.F. and Toulmin, S. (1975). Logic and the theory of mind: Formal, pragmatic and empirical considerations in a science of cognitive development. In J.K. Cole & W.J. Arnold (Eds.), Nebraska Symposium on Motivation (Vol. 23). Lincoln: University of Nebraska Press.

Feshbeck, S., Adelman, H., and Fuller, W. (1977). Prediction of reading and related academic problems. Journal of Educational Psychology, 69, 299-308.

Flavell, J.H. (1971). Stage-related properties of cognitive development. Cognitive Psychology, 2, 421-453.

Flavell, J.H. (1977). Cognitive Development, New Jersey: Prentice Hall.

Francis, H. (1973). Children's experience of reading and notions of units in language. British Journal of Educational. Psychology, 43, 17-23.

Gallagher, J.M. (1979). Problems in applying Piaget to reading, or letting the bird out of the cage. Journal of Education Boston University School of Education, 161, 72-86.

Goldstein, H. (1979). Some models for analysing longitudinal data on educational attainment. Journal of the Royal Statistical Society, 142, 407-442.

Golinkoff, R.M. (1976). A comparison of reading comprehension processes in good and poor comprehenders. Reading Research Quarterly, 2, 623-656.

Goodman, K.S. (1969). Analysis of oral reading miscues: applied psycholinguistics. Readind Research Quarterly. 2, 9-30.

Goodman, K.S. Behind the eye: What happens in reading. In K.S. Goodman and O.S. Niles (Eds.) (1970). Reading Process and Program. Illinois: National Council of Teachers of English. (Reprinted in H. Singer and R.B. Ruddell (Eds.), Theoretical Models and Processes of Reading, (2nd ed.). Newark, Delaware: I.R.A., 1976).

Hall, P. and Pacey, C. (1978). Paragraph Understanding Test. Sydney, N.S.W.: Department of Education.

Hooper, F.H. Toniolo, T.A. and Sipple, T.S. (1978). A longitudinal analysis of logical reasoning relationships: conservation and transitive inference. Developmental Psychology. 14, 674-682.

Howes, Polk, C.L., and Goldstein, D. (1980). Early reading and concrete operations. Journal of Psychology. 106, 111-116.

Huey, E.B. (1968). The Psychology and Pedagogy of Reading. Cambridge, Mass.: MIT Press. (Originally published, 1908).

Hull, C.H. and Nie, N.H. (1981). SPSS Update 7-9. New York: McGraw & Hill.

Inhelder, B., Sinclair, H. and Bovet, M. (1974). Learning and the Development of Cognition. London, Routledge.

Kagan, J. (1980). Perspectives on continuity. In O.G. Brim and J. Kagan (Eds.), Constancy and Change in Human Development. Cambridge, Mass.: Harvard University Press.

Kagan, J. and Moss, A.M. (1962). Birth to Maturity. New York: Wiley.

Katz, L. (1977). Reading ability and single-letter orthographic redundancy. Journal of Educational Psychology, 69, 653-659.

Kaufman, A.S. and Kaufman, N.L. (1972). Tests built from Piaget's and Gessell's tasks as predictors of first grade achievement. Child Development, 43, 521-535.

Keeton, A. (1977). Children's cognitive integration and memory processes for comprehending written sentences. Journal of Experimental Child Psychology, 23, 459-471.

Kerlinger, F.N. and Pedhazur, E.J. (9173). Multiple Regression in Behavioral Research. New York: Holt, Rinehart and Winston.

Kirk, S.A., McCarthy, J.J. and Kirk, W.E. (1968). Illinois Test of Psycholinguistic Abilities. University of Illinois Press.

Kolers, P. A. (1975). Pattern analysing in poor readers. Developmental Psychology, 2, 282-290.

La Berge, D. and Samuels, S. J. (1976). Toward a theory of automatic information processing in reading. In H. Singer, and R. Ruddell, (Eds.). Theoretical Models and Processes of Reading (2nd ed). Newark, Delaware. I.R.A., 458-576.

Larsen, G. (1977). Methodology in developmental psychology: as examination of research in Piagetian theory. Child Development. 48, 1160-1166.

Levin, J. (1973). Inducing Comprehension in Poor Readers: A test of a Recent Model. Journal of Educational Psychology. 65, 19-24.

Lunzer, E. A. (1970). Construction of a standardized battery of Piagetian tests to assess the development of effective intelligence. Research in Education, 3, 53-72.

Lunzer, E. L. and Dolan, T. (1977). Making Sense: Social Science, Research Council. Nottingham: University School of Education.

Lunzer, E. A., Dolan, T. and Wilkinson, J. E. (1976). The effectiveness of measures of operativity, language and short term memory in the prediction of reading and mathematical understanding. British Journal of Educational Psychology, 46, 295-305.

Marchbanks, G. and Levin, H. (1965). Cues by which children recognize words. Journal of Educational Psychology. 56, 57-61.

Marsh, H.W. and Hocevar, D. (1983). Confirmatory factor analysis of multitrait-multimethod matrices. Journal of Educational Measurement. 20, 231-248.

McCall, R. B. (1977). Challenges to a science of developmental psychology. Child Development, 48, 333-344.

McCall, R.B., Appelbaum, M.I. and Hogarty, P.S. (1973). Developmental Changes in mental performance. Monographs of the Society for Research in Child Development. 38 (3, Serial No. 150).

McCall, R.B., Eichorn, D.H. and Hogarty, P.S. (1977). Transitions in early mental development . Monographs of the Society for Research in Child Development. 42 (3 Serial No. 171).

McLeod, J. (1977). GAP Reading Comprehension, Form B3. Richmond, Victoria: Heinemann, 3rd Ed.

Meyer, B.J.F. (1984). Organisational aspects of text: Effects on reading comprehension and applications for the classroom. In J. Flood, Promoting reading comprehension. Newark, Delaware: International Reading Association.

Nesselroade, J.R., and Baltes, P.B. (1974). Adolescent personality development and historic change: 1970-1972. Monographs of the Society for Research in Child Development, 39 (1, Serial No. 154).

Nie, N. H., Hull, C. H., Jenkins, J. G., Steinbrenner, K., Bert, D. H. (1975). Statistical Package for the Social Sciences. (2nd Ed.), New York: McGraw Hill.

Piaget, J. (1967). The mental development of the child. In D. Elkind, (Ed.), Six Psychological Studies. New York: Random House.

Piaget, J. (1969). The Mechanisms of Perception. London: Routledge.

Ryan, E. B. (1979). Identifying and remediating failures in reading comprehension: toward an instructional approach for poor comprehenders. In T. G. Waller and G. E. Mackinnon, (Eds.), Reading Research: Advances in Theory and Practice (Vol. 2), New York: Academic Press.

Samuels, S.J., Dahl, P. and Archwamety, T. (1974). Effect of hypothesis test training on reading skill. Journal of Educational Psychology, 66, 835-844.

Scott, R. (1970). Perceptual skills, general intellectual ability race and later reading achievement. The Reading Teacher, 23, 660-669.

Serafica, F. C. and Sigel, I. E. (1970). Styles of categorization and reading disability. Journal of Reading Behaviour, 2,105-115.

Sinatra, R.C., Stahl-Gemake, J. and Berg, D.N. (1984). Improving reading comprehension of disabled readers through semantic mapping. The Reading Teacher, 38, 22-29.

Singer, H. and Donlan, D. (1982). Active comprehension: Problem-solving schema with question generation for comprehension of complex short stories. Reading Research Quarterly 17, 166-183.

Smith, F. (1971). Understanding Reading. New York: Holt Rinehart and Winston, Second ed. 1978, Third ed. 1982.

Steiner, R., Wiener, M. and Cromer, W. (1971). Comprehension training and identification for good and poor readers. Journal of Educational Psychology, 62, 506-513.

Stone, C.A. and Day, M.C. (1978). Levels of availability of a formal operational strategy. Child Development, 49, 1054-1065.

Sullivan, J.L. and Feldman, S. (1979). Multiple Indication: An Introduction. Beverly Hills, Sage.

Tomlinson-Keasey, C., Eisert, D.C., Kahle, D.C., Hardy-Brown, K., and Keasey, B. (1979). The structure of concrete operational thought. Child Development, 50, 1153-1163.

Uzgiris, I. (1976). Organisation of sensorimotor intelligence. In M. Lewis (Ed.),Origins of Intelligence. Infancy and Early Childhood, London: Wiley.

Valtin, R. (1978-1979). Dyslexia: deficit in or deficit in research? (Critical comments on the methodological and theoretical aspects of research on Legasthenil). Reading Research Quarterly, 14, 201-221.

Vurpillot, E. (1972). The Visual World of the Child. London: Allen and Unwin.

Waller, T. G. (1977). Think First, Read Later! Piagetian Prerequisites for Reading. Newark, Delaware: International Reading Association.

Watson, A. J. (1979). Multiple seriation and learning to read. Australian Journal of Education, 23, 171-180.

Watson, A. J. (1984). Cognition and units of print in early reading. In J. Downing and R. Valtin (Eds.) Language awareness and learning to read. New York: Springer.

White, S.H. (1965). Evidence for a hierarchical arrangement of learning processes. Advances in Child Development and Behavior, 2: 187-220.

White, S.H. (1970). Some general outlines of the matrix of developmental changes between five and seven years. Bulletin of the Orton Society, 20, 41-51.

Wohlwill, J. F. (1973). The Study of Behavioural Development. New York: Academic.

Wolfle, L. M. (1980). Strategies of path analysis. American Educational Research Journal, 17, 182-209.

Zaporozhets, A.V. (1969). In M. Cole and I. Maltzman (Eds.), Some of the psychological problems of sensory training in early childhood and the preschool period. A Handbook of Contemporary Soviet Psychology. New York: Basic.